

# Modelling spatial heteroskedasticity by volatility modulated moving averages

MICHELE NGUYEN and ALMUT E. D. VERAART

*Department of Mathematics, Imperial College London*

## Abstract

Spatial heteroskedasticity refers to stochastically changing variances and covariances in space. Such features have been observed in, for example, air pollution and vegetation data. We study how volatility modulated moving averages can model this by developing theory, simulation and statistical inference methods. For illustration, we also apply our procedure to sea surface temperature anomaly data from the International Research Institute for Climate and Society.

**Keywords:** moments-based inference, moving averages, stochastic simulation, stochastic volatility, spatial processes

## 1 Introduction

A classical assumption made when dealing with spatial data is that the variance is a constant and the covariance between measurements at two locations is a function of their distance apart. In practice, however, it has been observed that this does not hold for many data sets and accounting for spatial heteroskedasticity or spatial volatility has multiple benefits. The first benefit is the better representation of the data. In a recent paper, it was shown that including spatial volatility in road topography models better captures the hilliness features of the roads (Johannesson et al. 2016). This has implications on estimating the risk of vehicle damage and simulating fuel consumption. In some settings, the presence of spatial volatility can be also explained. For example, in a study of sulphur dioxide concentrations by Fuentes & Smith (2001), it was found that states which lie close to several coal power plants tend to have high variability in their readings. This was attributed to the dependence of the levels on the wind speed, the wind direction, as well as the atmospheric stability. A second benefit of modelling spatial volatility is the potential for improving prediction. This was seen by Huang et al. (2011) when they fitted a Gaussian process with volatility to vegetation and nitrate deposition data. In the case of agriculture yields, prediction intervals accounting for spatial volatility will be useful for insurance companies when they set crop insurance prices (Yan 2007).

Another way of using spatial volatility would be as an indicator of regime change. Such an approach has been taken in desertification and urban planning studies (Seekell & Dakos 2015, Getis 2015). In the first case, regions of high volatility demarcate the bare and the extensive vegetative cover; while in the second case, it is used to identify slum areas.

In this paper, we introduce stochastic volatility to the well-known Gaussian moving average (GMA) or process convolution model:

$$Y(\mathbf{x}) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi}) W(d\boldsymbol{\xi}), \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^d$  for some  $d \in \mathbb{N}$ ,  $g$  is a deterministic (kernel) function and  $W$  is the white noise on  $\mathbb{R}^d$  or a homogeneous standard Gaussian basis whose Lévy seed (which we shall define in Section 2) has mean 0 and variance 1. This results in the so-called volatility modulated moving average (VMMA):

$$Y(\mathbf{x}) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi}) \sigma(\boldsymbol{\xi}) W(d\boldsymbol{\xi}), \quad (2)$$

where  $\{\sigma^2(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \mathbb{R}^d\}$  is a stationary stochastic volatility field, independent of  $W$ . In Huang et al. (2011), the stochastic volatility is multiplied as a factor to the main spatial process; here, it appears as an integrand. As such,  $Y$  can sometimes be identified as a solution to a stochastic partial differential equation. Following similar arguments to those on page 559 of Bolin (2014), we find that  $Y$  can be viewed as a solution to:

$$(\kappa^2 - \Delta)^{\alpha/2} Y(\mathbf{x}) = \sigma(\mathbf{x}) \dot{W}(\mathbf{x}),$$

where  $\alpha > d/2$ ,  $\kappa > 0$ ,  $\Delta = \sum_{i=1}^d \partial^2 / \partial x_i^2$  is the Laplacian operator and  $\dot{W}$  is Gaussian white noise, when  $g$  is a Matérn kernel defined by:

$$g(\mathbf{x} - \boldsymbol{\xi}) = 2^{1-(\alpha-d)/2} (\kappa |\mathbf{x} - \boldsymbol{\xi}|)^{(\alpha-d)/2} K_{(\alpha-d)/2}(\kappa |\mathbf{x} - \boldsymbol{\xi}|) / [(4\pi)^{d/2} \Gamma(\alpha/2) \kappa^{\alpha-d}], \quad (3)$$

and  $K_{(\alpha-d)/2}$  is the modified Bessel function of the second kind.

VMMAAs can be seen as an extension of the Type G Lévy moving average (LMA) recently studied by Bolin (2014) and Wallin & Bolin (2015):

$$Y(\mathbf{x}) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi}) L(d\boldsymbol{\xi}), \quad (4)$$

where  $L$  is a (homogeneous) Type G Lévy basis. This means that the Lévy seed,  $L' \stackrel{d}{=} V^{1/2} Z$  where  $V$  is an infinitely divisible random variable and  $Z$  is a standard normal random variable independent of  $V$ . This is equivalent to restricting  $\sigma^2$  in the definition of our VMMA to be infinitely divisible and independent across locations.

The VMMA is also a special case of the volatility modulated mixed moving average studied by Veraart (2015):

$$\int_{\mathcal{X} \times \mathbb{R}^d} g(z, \mathbf{x} - \boldsymbol{\xi}) \sigma(\boldsymbol{\xi}) W(dz, d\boldsymbol{\xi}),$$

where  $\mathcal{X} \subset \mathbb{R}^k$  for some  $k \in \mathbb{N}$  and  $W$  is a more general Gaussian basis (Veraart 2015). It would be useful to study the simulation and inference procedures for the VMMA before moving on to this case where the parameters in the kernel function are randomised.

Convolution models such as GMAs and LMAs have been used in Geostatistics for designing spatial correlation structures. Apart from the classical stationary and isotropic correlation functions, other specifications can be made to construct non-stationary correlations with for example, locally varying geometric anisotropy (Fouedjio et al. 2016). In this paper, we show that VMMAAs give us the ability to model stationary and non-stationary correlations since the process is stationary but conditional on the volatility, non-stationary. Specifically, introducing  $\sigma$  to a GMA to form a VMMA retains the correlation constructed by our choice of  $g$  when we integrate or average over the realisations of  $\sigma$ . On the other hand, conditional on the realisation of  $\sigma$ , the VMMA has varying correlation structures over space.

Another usefulness of VMMAAs is that they enable us to model clustered extremes which are seen in many environmental data sets. This is because  $\sigma(\boldsymbol{\xi})$  acts like a local standard deviation for the driving noise  $W$  and  $\sigma^2(\boldsymbol{\xi})$  is modelled as a process with correlation. If clustered extremes are not suitable for the context, a good fit of the data to a VMMA could reveal missing covariates in the mean trend, location-dependent explanatory variables or areas where accurate measurements are hard to make.

**Outline** We begin in Section 2 by summarising the  $\mathcal{L}_0$  integration theory in Rajput & Rosinski (1989) and providing an integrability condition for the VMMA defined in (2). Next, we develop the theoretical properties of VMMAAs in Section 3. The main contributions of our research lie in the following two sections. In Section 4, we use discrete convolution ideas to design a simulation algorithm for VMMAAs. This is illustrated for a VMMA with a layered structure: the stochastic volatility field is an LMA and the kernels at both the VMMA and LMA levels are Gaussian. We provide a semi-explicit expression for the mean squared error and study cases where an explicit upper bound as well as its order of convergence can be obtained. In Section 5, we tackle the problem of inference for VMMAAs. We develop a two-step moments-matching estimation method which involves a moving window to obtain local variance estimates. Simulation experiments are conducted and the consistency of the estimators is proved under suitable double asymptotics. Next, we apply our

method to sea surface temperature anomaly data in Section 6 to illustrate benefits of using VMMA instead of GMA in this case. Finally, we conclude and discuss future steps for action in Section 7.

## 2 The $\mathcal{L}_0$ integration and integrability conditions

To construct the required stochastic integrals, we first need define our integrator: the homogeneous Lévy basis.

### 2.1 Homogeneous Lévy bases

Let  $\mathcal{B}(\mathbb{R}^d)$  denote the Borel  $\sigma$ -algebra on  $\mathbb{R}^d$  and  $\mathcal{B}_b(\mathbb{R}^d) = \{E \in \mathcal{B}(\mathbb{R}^d) : \text{Leb}(E) < \infty\}$  where  $\text{Leb}$  represents the Lebesgue measure. We work in the probability space  $(\Omega, \mathcal{F}, P)$ . To understand what a homogeneous Lévy basis is, we first define a Lévy basis (Barndorff-Nielsen et al. 2015, Sato 2007):

**Definition 1** (Lévy basis).

Let  $\{E_i : i \in \mathbb{N}\}$  be any sequence of the disjoint elements of  $\mathcal{B}_b(\mathbb{R}^d)$ . Suppose that  $L$  is a set of  $\mathbb{R}$ -valued random variables indexed by such sets, i.e.  $\{L(E) : E \in \mathcal{B}_b(\mathbb{R}^d)\}$  where, for  $\bigcup_{j=1}^{\infty} E_j \in \mathcal{B}_b(\mathbb{R}^d)$ ,  $L(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} L(E_j)$  almost surely. Then,  $L$  is called a random measure.

A random measure  $L$  is said to be a Lévy basis on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  if:

1. it is independently scattered:  $L(E_1), L(E_2), \dots$  are independent;
2. and it is infinitely divisible: the random vector  $\mathbf{L} = (L(B_1), \dots, L(B_m))$ , where  $B_1, \dots, B_m$  are elements of  $\mathcal{B}_b(\mathbb{R}^d)$ , is infinitely divisible. This means that there exists a law  $\mu_n$ , for any  $n \in \mathbb{N}$ , such that the law of  $\mathbf{L}$  can be written as  $\mu = \mu_n^{*n}$ , the  $n$ -fold convolution of  $\mu_n$  with itself.

Now, we specify a homogeneous Lévy basis as well as its so-called Lévy seed using the notion of a cumulant generating function (CGF). Note that the Lévy seed is important for defining the distributions of a Lévy basis and its associated moving average processes.

**Definition 2** (CGF, homogeneous Lévy basis and its seed).

The CGF of a random variable  $Z$ , which is denoted by  $C(\theta; Z)$ , is defined as the distinguished logarithm of its characteristic function, i.e.  $\log \mathbb{E}[\exp(i\theta Z)]$ .

Let  $L$  be a Lévy basis. Suppose that there exists a random variable  $L'$  such that  $C(\theta; L(E)) = C(\theta; L') \text{Leb}(E)$  for all  $E \in \mathcal{B}_b(S)$ . Then, we say that  $L$  is homogeneous and  $L'$  is its Lévy seed.

*Example 1.* The standard Gaussian basis in (2) is a homogeneous Lévy basis. In this case,  $W(E) \sim N(0, \text{Leb}(E))$  for any  $E \in \mathcal{B}_b(\mathbb{R}^d)$ .

*Example 2.* In this paper, we will also use the inverse Gaussian (IG) Lévy basis in our simulation studies. The parameterisation chosen is such that, for  $z > 0$ ,  $\delta > 0$  and  $\gamma > 0$ , the probability density function of a random variable  $Z$  with an  $IG(\delta, \gamma)$  distribution is:

$$f(z; \delta, \gamma) = \frac{\delta}{\sqrt{2\pi} z^3} \exp\left(\delta\gamma - \frac{1}{2}\left(\frac{\delta^2}{z} + \gamma^2 z\right)\right).$$

In this case, if  $L$  is an IG basis whose seed has an  $IG(\delta, \gamma)$  distribution,  $L(E) \sim IG(\delta \text{Leb}(E), \gamma)$  for any  $E \in \mathcal{B}_b(\mathbb{R}^d)$ .

## 2.2 Summary of the theory in Rajput & Rosinski (1989)

Since we only deal with homogeneous Lévy bases for the GMAs, LMAs and VMMA, we present the integration theory for this case. As usual, a stochastic integral is built up as a limit of those defined by so-called simple functions:

**Definition 3** (Simple function on  $\mathbb{R}^d$ , the stochastic integral of a simple function).

Consider  $\{y_j \in \mathbb{R} : j = 1, \dots, n\}$  and  $\{E_j : j = 1, \dots, n\}$ , a collection of disjoint sets of  $\mathcal{B}_b(\mathbb{R}^d)$ . Then,  $f(\mathbf{x}) = \sum_{j=1}^n y_j \mathbf{1}_{E_j}(\mathbf{x})$  is called a simple function on  $\mathbb{R}^d$  where  $\mathbf{1}_{E_j}(\mathbf{x}) = 1$  if  $\mathbf{x} \in E_j$  and 0 otherwise. The stochastic integral of  $f$  over  $A \in \mathcal{B}(\mathbb{R}^d)$  is defined as  $\int_A f(\boldsymbol{\xi}) L(d\boldsymbol{\xi}) = \sum_{j=1}^n y_j L(A \cap E_j)$ .

The stochastic integral of a measurable function is a simple extension of this:

**Definition 4** ( $L$ -integrability and the stochastic integral of a measurable function).

Let  $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  be a measurable function. Then,  $f$  is  $L$ -integrable if there exists a sequence of simple functions  $\{f_m\}$  such that:

- (i)  $f_m$  converges to  $f$  almost everywhere with respect to the Lebesgue measure;
- (ii) the sequence  $\{\int_A f_m(\boldsymbol{\xi}) L(d\boldsymbol{\xi})\}$  converges in probability for every  $A \in \mathcal{B}(\mathbb{R}^d)$ .

For an  $L$ -integrable function  $f$ , we define:

$$\int_A f(\boldsymbol{\xi}) L(d\boldsymbol{\xi}) = P - \lim_{m \rightarrow \infty} \int_A f_m(\boldsymbol{\xi}) L(d\boldsymbol{\xi}).$$

This construction is well-defined because the limit does not depend on the sequence  $\{f_m\}$ .

Theorem 2.7 in Rajput & Rosinski (1989) provides us with explicit conditions for  $\mathcal{L}_0$  integrability. When the kernel  $g$  is Lebesgue integrable and square-integrable, these conditions are satisfied for GMAs and LMAs whose Lévy bases have finite second moments. To use the  $\mathcal{L}_0$  theory to construct VMMA, we condition on the realisation of  $\sigma^2$  and treat  $g(\mathbf{x} - \boldsymbol{\xi})\sigma(\boldsymbol{\xi})$  as a measurable function. The condition required for a well-defined VMMA is then given by:

*Condition 1.*  $\int_{\mathbb{R}^d} g^2(\mathbf{x} - \boldsymbol{\xi}) \sigma^2(\boldsymbol{\xi}) < \infty$ .

As will be shown later, this quantity is equal to the conditional variance of  $Y$  at  $\mathbf{x}$ . This is easy to see that the condition holds whenever  $g$  is square-integrable and  $\sigma^2$  takes finite values.

*Example 3* (Two-tiered model). An example of a well-defined VMMA is the so-called two-tiered model:

$$\left. \begin{aligned} Y(\mathbf{x}) &= \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi}) \sigma(\boldsymbol{\xi}) W(d\boldsymbol{\xi}), \\ \text{where } \sigma^2(\boldsymbol{\xi}) &= \int_{\mathbb{R}^d} h(\boldsymbol{\xi} - \mathbf{u}) L(d\mathbf{u}). \end{aligned} \right\} \quad (5)$$

Here,  $g$  and  $h$  are Lebesgue integrable and square-integrable kernel functions, and  $L$  is a subordinator with finite second moments (so that  $\sigma^2$  is well-defined). As before,  $W$  is a homogeneous standard Gaussian basis independent of  $\sigma^2$ .

Note that we model  $\sigma^2$  directly. In comparison, treating the volatility as a multiplicative factor as is done in Huang et al. (2011) is synonymous with modelling the conditional variance  $\int_{\mathbb{R}^d} g^2(\mathbf{x} - \boldsymbol{\xi}) \sigma^2(\boldsymbol{\xi}) d\boldsymbol{\xi}$ . In this case, we have chosen to model  $\sigma^2$  as an LMA because it is convenient for deriving the second order distributional properties of  $\sigma^2$ . When the VMMA can be viewed as a solution to an SPDE,  $\sigma^2$  could correspond to a process of special interest. For example, it could represent the cumulative effect which the wind speed, direction and atmospheric stability has on the spatial heteroskedasticity of air pollution. Thus, in Section 5, we are particularly interested in estimating the parameters of  $\sigma^2$  that determine its variance and correlation structure.

*Example 4.* As our main illustration example, we will use a more precise model:

$$\left. \begin{aligned} Y(\mathbf{x}) &= \int_{\mathbb{R}^2} \frac{\lambda}{\pi} \exp\left(-\lambda(\mathbf{x} - \boldsymbol{\xi})^T(\mathbf{x} - \boldsymbol{\xi})\right) \sigma(\boldsymbol{\xi}) W(d\boldsymbol{\xi}), \\ \text{where } \sigma^2(\boldsymbol{\xi}) &= \int_{\mathbb{R}^2} \frac{\eta}{\pi} \exp\left(-\eta(\boldsymbol{\xi} - \mathbf{u})^T(\boldsymbol{\xi} - \mathbf{u})\right) L(d\mathbf{u}). \end{aligned} \right\} \quad (6)$$

The stochastic volatility field is an LMA with the same kernel structure as the VMMA itself but a different rate parameter. Here, we choose Gaussian kernels with rate parameters  $\lambda, \eta > 0$ . As mentioned in Higdon (1998), these kernels are computationally convenient and are supported by physical ocean dynamics.

### 3 Theoretical properties of VMMA

In this section, we prove several distributional properties of VMMA including stationarity, cumulant and covariance structures. These will be useful for the estimation method which we develop later.

#### 3.1 Marginal distribution

##### 3.1.1 Conditional distribution and cumulants

Let  $\mathcal{F}^\sigma$  be the  $\sigma$ -algebra generated by the stochastic volatility  $\sigma^2$ . As we have assumed that  $\sigma$  and  $W$  are independent,  $Y(\mathbf{x})|\mathcal{F}^\sigma \sim N(0, \sigma_I^2(\mathbf{x}))$ , where  $\sigma_I^2(\mathbf{x}) = \int_{\mathbb{R}^d} g^2(\mathbf{x} - \boldsymbol{\xi}) \sigma^2(\boldsymbol{\xi}) d\boldsymbol{\xi}$  denotes the conditional variance.

Recall that the cumulants of the VMMA,  $\kappa_l(Y(\mathbf{x}))$ , are defined through its CGF. That is,  $C(\theta; Y) = \sum_{l=1}^{\infty} \kappa_l(Y(\mathbf{x})) \frac{(i\theta)^l}{l!}$ . In this case, since we have a Gaussian distribution, the conditional cumulants are  $\kappa_1^\sigma = \mathbb{E}[Y(\mathbf{x})|\mathcal{F}^\sigma] = 0$ ,  $\kappa_2^\sigma = \text{Var}[Y(\mathbf{x})|\mathcal{F}^\sigma] = \sigma_I^2(\mathbf{x})$  and  $\kappa_l^\sigma = 0$  for  $l \geq 3$ .

##### 3.1.2 Unconditional distribution and cumulants

From the first conditional cumulant, we get  $\kappa_1 = \mathbb{E}[Y(\mathbf{x})] = \mathbb{E}[\kappa_1^\sigma] = 0$ . So,  $\kappa_2 = \mathbb{E}[Y^2(\mathbf{x})] = \mathbb{E}[\sigma_I^2(\mathbf{x})]$ . Higher order unconditional cumulants can be calculated in similar ways; however, beyond the third cumulant, they are typically not be equal to 0 unlike their conditional counterparts. This is because the unconditional marginal distribution of a VMMA is generally not Gaussian.

*Example 5.* For Model (6), we have  $\kappa_1 = 0$ . Let  $\mathbb{E}[L'] = a < \infty$ . The next three cumulants are given by:

$$\begin{aligned} \kappa_2^\sigma = \sigma_I^2(\mathbf{x}) &\Rightarrow \kappa_2 = \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) \mathbb{E}[\sigma^2(\boldsymbol{\xi})] d\boldsymbol{\xi} = a \int_{\mathbb{R}^2} \frac{\lambda^2}{\pi^2} \exp\left(-2\lambda(\mathbf{x} - \boldsymbol{\xi})^T(\mathbf{x} - \boldsymbol{\xi})\right) d\boldsymbol{\xi} = \frac{a\lambda}{2\pi}. \\ \kappa_3^\sigma = 0 &\Rightarrow \mathbb{E}[Y^3(\mathbf{x})|\mathcal{F}^\sigma] - 3\mathbb{E}[Y^2(\mathbf{x})|\mathcal{F}^\sigma] \mathbb{E}[Y(\mathbf{x})|\mathcal{F}^\sigma] + 2\mathbb{E}^3[Y(\mathbf{x})|\mathcal{F}^\sigma] = 0. \Rightarrow \mathbb{E}[Y^3(\mathbf{x})] = \mathbb{E}[Y^3(\mathbf{x})|\mathcal{F}^\sigma] = 0. \\ &\Rightarrow \kappa_3 = \mathbb{E}[Y^3(\mathbf{x})] - 3\mathbb{E}[Y^2(\mathbf{x})] \mathbb{E}[Y(\mathbf{x})] + 2\mathbb{E}^3[Y(\mathbf{x})] = 0. \\ \kappa_4^\sigma = 0 &\Rightarrow \mathbb{E}[Y^4(\mathbf{x})|\mathcal{F}^\sigma] - 4\mathbb{E}[Y^3(\mathbf{x})|\mathcal{F}^\sigma] \mathbb{E}[Y(\mathbf{x})|\mathcal{F}^\sigma] - 3\mathbb{E}^2[Y^2(\mathbf{x})|\mathcal{F}^\sigma] + 12\mathbb{E}[Y^2(\mathbf{x})|\mathcal{F}^\sigma] \mathbb{E}^2[Y(\mathbf{x})|\mathcal{F}^\sigma] \\ &\quad - 6\mathbb{E}^4[Y(\mathbf{x})|\mathcal{F}^\sigma] = 0. \Rightarrow \mathbb{E}[Y^4(\mathbf{x})|\mathcal{F}^\sigma] = 3\mathbb{E}^2[Y^2(\mathbf{x})|\mathcal{F}^\sigma] = 3(\sigma_I^2(\mathbf{x}))^2. \\ &\Rightarrow \kappa_4 = \mathbb{E}[Y^4(\mathbf{x})] - 3\mathbb{E}^2[Y^2(\mathbf{x})] = 3 \text{Var}[\sigma_I^2(\mathbf{x})] = \frac{3b\lambda^3\eta}{4\pi^3(2\lambda + \eta)}, \text{ where } b = \text{Var}(L') \text{ and we have used (17).} \end{aligned}$$

## 3.2 Finite dimensional distributions

### 3.2.1 Conditional joint distribution and correlation structure

Now, we consider joint distributions of the process at different locations. This is characterised by the joint cumulant generating function (JCGF). To compute this for the VMMA, we introduce the concept of a generalised cumulant functional. This is a spatial extension of the concept given in Barndorff-Nielsen et al. (2015):

**Definition 5** (Generalised cumulant functional).

Let  $Y = \{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^d\}$  denote a stochastic process in  $\mathbb{R}^d$ , and let  $v$  denote any non-random measure such that  $v(Y) = \int_{\mathbb{R}^d} Y(\mathbf{x})v(d\mathbf{x})$ , where the integral exists almost surely. The generalised cumulant functional (GCF) of  $Y$  with respect to  $v$  is given by:  $C(\theta; v(Y)) = \log \mathbb{E}[\exp(i\theta v(Y))]$ .

To compute the JCGF of a VMMA,  $Y$ , we first condition on  $\sigma^2(\boldsymbol{\xi})$  and obtain the conditional GCF:

**Theorem 1.** *Let  $Y(\mathbf{x})$  be a VMMA defined by (2). Assume that for all  $\boldsymbol{\xi} \in \mathbb{R}^d$ ,  $h(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi})\sigma(\boldsymbol{\xi})v(d\mathbf{x}) < \infty$ , and that  $h(\boldsymbol{\xi})$  is integrable with respect to the Gaussian basis  $W$ . Then, with  $W'$  denoting the seed of  $W$ , the GCF of  $Y|\sigma^2$  with respect to  $v$  can be expressed as:*

$$C(\theta; v(Y)|\mathcal{F}^\sigma) = \int_{\mathbb{R}^d} C(\theta h(\boldsymbol{\xi}); W')d\boldsymbol{\xi} = -\frac{1}{2}\theta^2 \int_{\mathbb{R}^d} h^2(\boldsymbol{\xi})d\boldsymbol{\xi}.$$

*Proof.* This is analogous to the proof for Proposition 5 in Barndorff-Nielsen et al. (2015) with  $h$  being defined differently to account for our definition of  $Y(\mathbf{x})$ , and with the Lévy basis restricted to be standard Gaussian.  $\square$

Now, we use  $v(d\mathbf{x}) = \sum_{j=1}^n \theta_j \delta_{\mathbf{x}_j}(d\mathbf{x})$  so that  $C(\theta; v(Y)|\mathcal{F}^\sigma)$  is the JCGF of  $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)|\mathcal{F}^\sigma$ :

**Corollary 1.** *Let  $\mathbf{x}_1, \dots, \mathbf{x}_n$  be different locations in  $\mathbb{R}^d$ . The JCGF of  $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)|\mathcal{F}^\sigma$  is given by:*

$$\log \mathbb{E} \left[ \exp \left( i \sum_{j=1}^n \tilde{\theta}_j Y(\mathbf{x}_j) \right) | \mathcal{F}^\sigma \right] = -\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \tilde{\theta}_j \tilde{\theta}_k \int_{\mathbb{R}^d} g(\mathbf{x}_j - \boldsymbol{\xi})g(\mathbf{x}_k - \boldsymbol{\xi})\sigma^2(\boldsymbol{\xi})d\boldsymbol{\xi}.$$

This means that  $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)|\mathcal{F}^\sigma \sim N_n(\mathbf{0}, \Sigma)$  where  $\Sigma_{jk} = \int_{\mathbb{R}^d} g(\mathbf{x}_j - \boldsymbol{\xi})g(\mathbf{x}_k - \boldsymbol{\xi})\sigma^2(\boldsymbol{\xi})d\boldsymbol{\xi}$  for  $j, k = 1, \dots, n$ .

*Proof.* We first compute  $h(\boldsymbol{\xi})$ :

$$h(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi})\sigma(\boldsymbol{\xi})v(d\mathbf{x}) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi})\sigma(\boldsymbol{\xi}) \sum_{j=1}^n \theta_j \delta_{\mathbf{x}_j}(d\mathbf{x}) = \sum_{j=1}^n \theta_j g(\mathbf{x}_j - \boldsymbol{\xi})\sigma(\boldsymbol{\xi}).$$

With  $\tilde{\theta}_j = \theta_j$ , the JCGF of  $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)|\mathcal{F}^\sigma$  is given by:

$$\log \mathbb{E} \left[ \exp \left( i \sum_{j=1}^n \tilde{\theta}_j Y(\mathbf{x}_j) \right) | \mathcal{F}^\sigma \right] = -\frac{1}{2}\theta^2 \int_{\mathbb{R}^d} h^2(\boldsymbol{\xi})d\boldsymbol{\xi} = -\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \tilde{\theta}_j \tilde{\theta}_k \int_{\mathbb{R}^d} g(\mathbf{x}_j - \boldsymbol{\xi})g(\mathbf{x}_k - \boldsymbol{\xi})\sigma^2(\boldsymbol{\xi})d\boldsymbol{\xi}. \quad (7)$$

This corresponds to a multivariate normal distribution with the parameters stated in the Theorem.  $\square$

From the multivariate normal distribution, we can infer the covariance and correlation structures of our VMMA. Let  $\mathbf{x}$  and  $\mathbf{x}^*$  be two different locations in  $\mathbb{R}^d$ , then  $\text{Cov}(Y(\mathbf{x}), Y(\mathbf{x}^*)|\mathcal{F}^\sigma) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi})g(\mathbf{x}^* - \boldsymbol{\xi})\sigma^2(\boldsymbol{\xi})d\boldsymbol{\xi}$  and:

$$\text{Cov}(Y(\mathbf{x}), Y(\mathbf{x}^*)) = \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi})g(\mathbf{x}^* - \boldsymbol{\xi})\mathbb{E}[\sigma^2(\boldsymbol{\xi})]d\boldsymbol{\xi} = \mathbb{E}[\sigma^2(\mathbf{0})] \int_{\mathbb{R}^d} g(\mathbf{w})g(\mathbf{x}^* - \mathbf{x} + \mathbf{w})d\mathbf{w},$$

since  $\sigma^2(\boldsymbol{\xi})$  is stationary and where  $\mathbf{w} = \mathbf{x} - \boldsymbol{\xi}$ . As this is a function of the location difference and not the locations themselves,  $Y(\mathbf{x})$  has second-order stationarity. From the covariance function, we also find that the correlation structure does not depend on the stochastic volatility:

$$\text{Corr}(Y(\mathbf{x}), Y(\mathbf{x}^*)) = \frac{\int_{\mathbb{R}^d} g(\mathbf{w})g(\mathbf{x}^* - \mathbf{x} + \mathbf{w})d\mathbf{w}}{\int_{\mathbb{R}^d} g^2(\mathbf{w})d\mathbf{w}}.$$

An LMA with zero mean and the same kernel will also have this correlation structure. The effects of the stochastic volatility in the VMMA, and hence a difference from the GMA and the LMA, lies in the higher order correlations.

*Example 6.* For Model (6), we have  $g(\mathbf{w}) = \frac{\lambda}{\pi} \exp(-\lambda \mathbf{w}^T \mathbf{w})$ . By completing the squares:

$$\begin{aligned} \int_{\mathbb{R}^2} g(\mathbf{w})g(\mathbf{x}^* - \mathbf{x} + \mathbf{w})d\mathbf{w} &= \frac{\lambda^2}{\pi^2} \int_{\mathbb{R}^2} \exp(-\lambda \mathbf{w}^T \mathbf{w}) \exp(-\lambda (\mathbf{x}^* - \mathbf{x} + \mathbf{w})^T (\mathbf{x}^* - \mathbf{x} + \mathbf{w})) d\mathbf{w} \\ &= \frac{\lambda^2}{\pi^2} \int_{\mathbb{R}} \exp(-\lambda [w_1^2 + (w_1 - x_1 + x_1^*)^2]) dw_1 \int_{\mathbb{R}} \exp(-\lambda [w_2^2 + (w_2 - x_2 + x_2^*)^2]) dw_2 \\ &= \frac{\lambda}{2\pi} \exp\left(-\frac{\lambda (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*)}{2}\right). \\ \Rightarrow \text{Corr}(Y(\mathbf{x}), Y(\mathbf{x}^*)) &= \exp\left(-\frac{\lambda (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*)}{2}\right). \end{aligned}$$

**Corollary 2.** Let  $Y(\mathbf{x})$  be a VMMA and  $\mathbf{x}, \mathbf{x}^*$  denote two arbitrary locations in  $\mathbb{R}^d$ . Then:

$$\begin{aligned} \text{Cov}(Y^2(\mathbf{x}), Y^2(\mathbf{x}^*) | \mathcal{F}^\sigma) &= \mathbb{E}[Y^2(\mathbf{x})Y^2(\mathbf{x}^*) | \mathcal{F}^\sigma] - \sigma_I^2(\mathbf{x})\sigma_I^2(\mathbf{x}^*) = 2 \left( \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi})g(\mathbf{x}^* - \boldsymbol{\xi})\sigma^2(\boldsymbol{\xi})d\boldsymbol{\xi} \right)^2, \\ \text{and } \text{Cov}(Y^2(\mathbf{x}), Y^2(\mathbf{x}^*)) &= \mathbb{E}[\text{Cov}(Y^2(\mathbf{x}), Y^2(\mathbf{x}^*) | \mathcal{F}^\sigma)] + \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x}^*)) \\ &= 2\mathbb{E} \left[ \left( \int_{\mathbb{R}^d} g(\mathbf{x} - \boldsymbol{\xi})g(\mathbf{x}^* - \boldsymbol{\xi})\sigma^2(\boldsymbol{\xi})d\boldsymbol{\xi} \right)^2 \right] + \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x}^*)). \end{aligned}$$

*Proof.* We calculate  $\mathbb{E}[Y^2(\mathbf{x})Y^2(\mathbf{x}^*) | \mathcal{F}^\sigma]$  by setting  $n = 2$ , differentiating the conditional JCGF in (7) with respect to  $\tilde{\theta}_1$  and  $\tilde{\theta}_2$  twice each, and setting these to be equal to 0. The rest follows easily.  $\square$

*Example 7.* For Model (6):

$$\text{Cov}(Y^2(\mathbf{x}), Y^2(\mathbf{x}^*)) = A \exp(-\lambda (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*)) + B \exp\left(\frac{-\lambda \eta}{2\lambda + \eta} (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*)\right),$$

where  $A = (b\lambda\eta + a^2(2\lambda + \eta)\pi)\lambda^2(2\pi^3(2\lambda + \eta))^{-1}$  and  $B = b\lambda^3\eta(4\pi^3(2\lambda + \eta))^{-1}$ . The details of the computation can be found in the Appendix.

### 3.3 Unconditional joint distribution and stationarity

By exponentiating the expression in Corollary 1 for the conditional JCGF and taking expectations with respect to  $\sigma^2$ , we obtain the unconditional joint characteristic function (JCF) of  $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)$ :

$$\mathbb{E} \left[ \exp \left( i \sum_{j=1}^n \tilde{\theta}_j Y(\mathbf{x}_j) \right) \right] = \mathbb{E} \left[ \exp \left( -\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \tilde{\theta}_j \tilde{\theta}_k \int_{\mathbb{R}^d} g(\mathbf{x}_j - \boldsymbol{\xi})g(\mathbf{x}_k - \boldsymbol{\xi})\sigma^2(\boldsymbol{\xi})d\boldsymbol{\xi} \right) \right].$$

For specific  $g$  and  $\sigma^2$ , this can be expressed as a function of the location differences since  $Y$  is stationary:

**Theorem 2.** *Let  $Y(\mathbf{x})$  be a VMMA. Then,  $Y(\mathbf{x})$  is a stationary process in  $\mathbb{R}^d$ .*

*Proof.* We present the proof for  $\mathbf{x}$  in  $\mathbb{R}$ . The case for general  $\mathbb{R}^d$  follows analogously with more involved notation. Let  $x_0 < \dots < x_{n-1}$  denote  $n$  arbitrary locations in  $\mathbb{R}$ . We show that for any  $u \in \mathbb{R}$ , the JCF of  $Y(x_0 + u), \dots, Y(x_{n-1} + u)$  is the same as that of  $Y(x_0), \dots, Y(x_{n-1})$ :

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( i \sum_{j=0}^{n-1} \theta_j \int_{\mathbb{R}} g(x_j + u - \xi) \sigma(\xi) W(d\xi) \right) \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \exp \left( i \sum_{j=0}^{n-1} \theta_j \int_{\mathbb{R}} g(x_j + u - \xi) \sigma(\xi) W(d\xi) \right) \middle| \mathcal{F}^\sigma \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \exp \left( i \sum_{j=0}^{n-1} \theta_j \lim_{p \rightarrow \infty} \sum_{I=2}^{k_p-1} g(x_j + u - y_I) \sigma(y_I) W((y_I - (y_I - y_{I-1})/2, y_I + (y_{I+1} - y_I)/2)) \right) \middle| \mathcal{F}^\sigma \right] \right], \end{aligned} \quad (8)$$

where we follow the  $\mathcal{L}_0$  integration theory and use an approximating sequence for  $g(x_j + u - \xi)\sigma(\xi)$ :

$$f_p(\xi) = \sum_{I=2}^{k_p-1} g(x_j + u - y_I) \sigma(y_I) \mathbf{1}_{(y_I - (y_I - y_{I-1})/2, y_I + (y_{I+1} - y_I)/2]}(\xi),$$

where  $(y_{k_p})$  is a sequence of partitions such that:

$$-\infty < y_1 < y_2 < \dots < y_{k_p} < \infty,$$

$k_p \rightarrow \infty$  and  $\max_{I \in \{2, \dots, k_p\}} (y_I - y_{I-1}) \rightarrow 0$  as  $p \rightarrow \infty$ . Note that we have assumed that  $g$  takes finite values so that this approximating sequence can be evaluated. If  $g$  has a finite number of singularities, for example in the case of the Matérn kernel (3) with  $(\alpha - d)/2 \in (-1/2, 0)$ , shifts can be made to the evaluation points to avoid these.

Define another sequence of partitions  $(z_{k_p}) = (y_{k_p} - u)$ . By changing the order of taking limits and sums, and using  $h(z_I) = \sum_{j=0}^{n-1} \theta_j g(x_j - z_I) \sigma(z_I + u)$ , the JCF can be written as:

$$\begin{aligned} & \lim_{p \rightarrow \infty} \mathbb{E} \left[ \mathbb{E} \left[ \exp \left( i \sum_{I=2}^{k_p-1} h(z_I) W((z_I + u - (z_I - z_{I-1})/2, z_I + u + (z_{I+1} - z_I)/2)) \right) \middle| \mathcal{F}^\sigma \right] \right] \\ &= \lim_{p \rightarrow \infty} \mathbb{E} \left[ \prod_{I=2}^{k_p-1} \mathbb{E} [\exp(ih(z_I) W((z_I + u - (z_I - z_{I-1})/2, z_I + u + (z_{I+1} - z_I)/2))] \middle| \mathcal{F}^\sigma \right] \end{aligned} \quad (9)$$

$$= \lim_{p \rightarrow \infty} \mathbb{E} \left[ \prod_{I=2}^{k_p-1} \exp \left( -\frac{z_I + (z_{I+1} - z_I)/2 - z_I + (z_I - z_{I-1})/2}{2} h^2(z_I) \right) \right] \quad (10)$$

where (9) and (10) hold because  $W$  is independently scattered and homogeneous standard Gaussian.

Recall that  $\sigma^2$  is stationary. Since the term inside the expectation in (10) is a Borel transformation of  $(\sigma(z_1 + u), \dots, \sigma(z_{k_p} + u))$ , it has the same distribution and expectation as a similar expression with  $(\sigma(z_1), \dots, \sigma(z_{k_p}))$  instead. Since (10) no longer depends on  $u$ , we conclude that  $Y$  is stationary.  $\square$



## 4 Simulation

### 4.1 A discrete convolution algorithm

We focus on cases in  $\mathbb{R}^2$  where the kernel function in the VMMA takes finite values. Let  $\mathbf{x} = \{(x_1^i, x_2^j) = (x_1^* + i\Delta, x_2^* + j\Delta) : i, j = -p, \dots, N + p - 1\}$  be our simulation grid where  $(x_1^*, x_2^*)$  is the starting point,  $\Delta$  is the grid size,  $p \in \mathbb{N}$  is a kernel truncation parameter and  $N \in \mathbb{N}$  is the number of coordinates in each spatial axis so that the final sample size is  $N^2$ . By discretizing the stochastic integral in (2), we can view a VMMA as a filtered process where  $g(\mathbf{x} - \boldsymbol{\xi})$  is the kernel or filter and  $\sigma(\boldsymbol{\xi})W^*(\boldsymbol{\xi})$  is the signal. Here,  $W^*(\boldsymbol{\xi}) \stackrel{d}{=} N(0, \Delta^2)$  is a random variable representing the Gaussian noise over the grid square centered at  $\boldsymbol{\xi}$ . These  $W^*$ s are independent across locations. In practice, this means that we approximate our VMMA by:

$$Y(x_1^i, x_2^j) \approx \sum_{I=i-p}^{i+p} \sum_{J=j-p}^{j+p} g(x_1^i - x_1^I, x_2^j - x_2^J) \sigma(x_1^I, x_2^J) W^*(x_1^I, x_2^J), \quad (11)$$

for  $i, j = 0, \dots, N - 1$ . If  $g$  is square-integrable,  $g(\mathbf{x} - \boldsymbol{\xi})$  typically decreases very fast to 0 as  $|\mathbf{x} - \boldsymbol{\xi}| \rightarrow 0$ . Thus, only small errors are incurred by truncating the kernel. Following the  $\mathcal{L}_0$  integration theory and the proof of Theorem 2, (11) can also be viewed as an approximation of  $Y$  when  $g$  is approximated using a particular equispaced partition and the increments of  $W$  are replaced by independent, identically distributed random variables  $W^*$ .

Suppose that we know the values of  $\sigma$  over  $\mathbf{x}$ , then we can generate values for  $\mathbf{Y} = \{Y(x_1^i, x_2^j) : i, j = 0, \dots, N - 1\}$  using (11). To begin, we create a  $(2p + 1) \times (2p + 1)$  kernel matrix  $K$  as follows:

$$K = \begin{pmatrix} g(p\Delta, -p\Delta) & \dots & g(-p\Delta, -p\Delta) \\ \vdots & \vdots & \vdots \\ \dots & g(0, 0) & \dots \\ \vdots & \vdots & \vdots \\ g(p\Delta, p\Delta) & \dots & g(-p\Delta, p\Delta) \end{pmatrix}. \quad (12)$$

Then, we generate  $\{W^*(x_1^i, x_2^j)\}$  for  $i, j = -p, \dots, N + p - 1$ . We multiply each of these  $W^*$ s to their corresponding  $\sigma$  values and create a  $(N + 2p) \times (N + 2p)$  signal matrix:

$$\sigma W^* = \begin{pmatrix} \sigma(x_1^{-p}, x_2^{N+p-1}) W^*(x_1^{-p}, x_2^{N+p-1}) & \dots & \sigma(x_1^{N+p-1}, x_2^{N+p-1}) W^*(x_1^{N+p-1}, x_2^{N+p-1}) \\ \vdots & \vdots & \vdots \\ \sigma(x_1^{-p}, x_2^{-p}) W^*(x_1^{-p}, x_2^{-p}) & \dots & \sigma(x_1^{N+p-1}, x_2^{-p}) W^*(x_1^{N+p-1}, x_2^{-p}) \end{pmatrix}. \quad (13)$$

From (11), we see that a matrix of VMMA values  $Y = \{Y(x_1^i, x_2^j) : i, j = 0, \dots, N - 1\}$  can be obtained through a filtering of  $\sigma W^*$  by  $K$ . To compute this efficiently, we can use the convolution theorem and fast Fourier transform (fft) schemes which are available in software such as R. As summarised in Algorithm 1, this involves zero-padding  $K$  and  $\sigma W^*$  to the size  $(N + 4p) \times (N + 4p)$ , taking the Fourier transforms of the resulting matrices and cropping the inverse Fourier transform of their element-wise product.

For Model (6), we can compute  $\sigma^2(\boldsymbol{\xi})$  first by replacing  $\sigma W^*$  in Algorithm 1 by a  $(N + 2p) \times (N + 2p)$  matrix of generated values for the Lévy noise over each grid square. Figure 1(a) shows a simulated stochastic volatility layer over the region  $\mathbf{x} = [-1.5, 11.5] \times [-1.5, 11.5]$  where  $\eta = 4$  and  $L$  is an IG basis whose seed has mean  $a = 1$  and variance  $b = 2$ . Here, we have chosen  $N = 261$ ,  $\Delta = 0.05$  and  $p = 30$ . After taking square roots of the volatility values and multiplying the results with the Gaussian realisations, we can use Algorithm 1 again to compute  $Y$ . Assuming that the same kernel truncation parameter  $p$  is used, this results in  $(261 - 2p)^2 = 201^2$  values. This scheme takes about two seconds to generate one VMMA data set using a PC with characteristics: Intel® Core™i7-3770 CPU Processor @ 3.40GHz; 8GB of RAM; Windows 8.1 64-bit. An example of a simulated realisation is shown in Figure 1(b).

**Algorithm 1** Discrete convolution via Fourier transforms

---

```

1:  $M1 \leftarrow \text{matrix}(0, N + 4p, N + 4p)$  ▷ We create a  $(N + 4p) \times (N + 4p)$  matrix of 0s.
2:  $M1[1 : (2p + 1), 1 : (2p + 1)] \leftarrow K$  ▷ We insert  $K$  into the top left corner of  $M1$ .
3:  $M2 \leftarrow \text{matrix}(0, N + 4p, N + 4p)$  ▷ We create another  $(N + 4p) \times (N + 4p)$  matrix of 0s.
4:  $M2[1 : (N + 2p), 1 : (N + 2p)] \leftarrow \sigma W^*$  ▷ We insert  $\sigma W^*$  into the top left corner of  $M2$ .
5:  $FM1 \leftarrow \text{fft}(M1, \text{inverse} = \text{FALSE})$  ▷ We compute the forward fft of  $M1$ .
6:  $FM2 \leftarrow \text{fft}(M2, \text{inverse} = \text{FALSE})$  ▷ We compute the forward fft of  $M2$ .
7:  $FM \leftarrow FM1 * FM2$  ▷ We multiply  $FM1$  and  $FM2$  element-wise.
8:  $Y \leftarrow \text{Re}(\text{fft}(FM, \text{inverse} = \text{TRUE}) / (N + 4p)^2)$  ▷ We take the real part of the inverse Fourier transform of  $FM$ .
9:  $Y \leftarrow Y[(2p + 1) : (N + 2p), (2p + 1) : (N + 2p)]$  ▷ We crop the matrix  $Y$  to obtain the final filtered process.

```

---

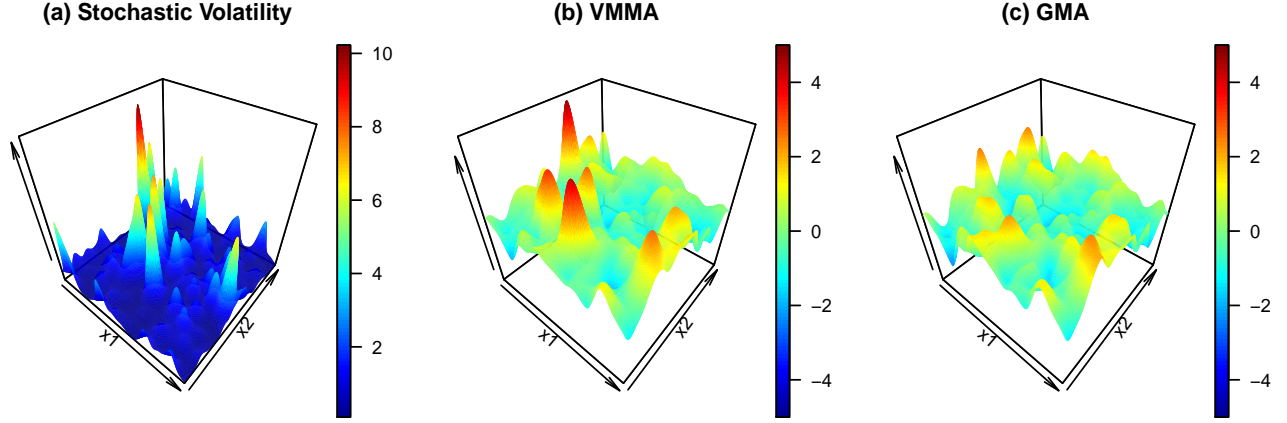


Figure 1: Sample paths related to Model (6): (a) the stochastic volatility with  $\eta = 4$ ; (b) the VMMA with  $\lambda = 4$  and IG basis whose seed has mean  $a = 1$  and variance  $b = 2$ ; (c) the GMA with the same kernel structure as the VMMA. The same realisation of the Gaussian driving noise is used for the VMMA and GMA to facilitate comparison.

Such a simulation scheme can also be used for generating data for GMAs. In Figures 1(c), we show the realisation from a GMA with the same underlying Gaussian noise as the VMMA in Plot (b). The VMMA exhibits clustered extremes where the values of its stochastic volatility are high in Plot (a). This in turn has the effect of smoothing the VMMA surface as seen from the contours in Figure 2(d). The first column of plots in Figure 2 correspond to the same realisations of the stochastic volatility, VMMA and GMA as those in Figure 1.

## 4.2 Mean-square error and its upper bound

We have a two-step discrete convolution (TSDC) simulation algorithm for the two-tiered VMMA defined in (5). Let  $Y$  be such a VMMA and  $Z = \{Z(\mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^2}$  be its TSDC approximation. Then, we can write  $Z(\mathbf{x}) = \int_{\mathbb{R}^2} g_{\Delta}(\mathbf{x}, \boldsymbol{\xi}) \sigma_{\Delta}(\boldsymbol{\xi}) W(d\boldsymbol{\xi})$ , where:

$$\begin{aligned}
g_{\Delta}(\mathbf{x}, \boldsymbol{\xi}) &= \sum_{i=-p}^p \sum_{j=-p}^p \mathbf{1}_{[x_1+i\Delta-\frac{\Delta}{2}, x_1+i\Delta+\frac{\Delta}{2})}(s_1) \mathbf{1}_{[x_2+j\Delta-\frac{\Delta}{2}, x_2+j\Delta+\frac{\Delta}{2})}(s_2) g(i\Delta, j\Delta), \\
\sigma_{\Delta}^2(\boldsymbol{\xi}) &= \int_{\mathbb{R}^2} h_{\Delta}(\boldsymbol{\xi}, \mathbf{u}) L(d\mathbf{u}), \\
h_{\Delta}(\boldsymbol{\xi}, \mathbf{u}) &= \sum_{i=-\tilde{p}}^{\tilde{p}} \sum_{j=-\tilde{p}}^{\tilde{p}} \mathbf{1}_{[\xi_1+i\Delta-\frac{\Delta}{2}, \xi_1+i\Delta+\frac{\Delta}{2})}(u_1) \mathbf{1}_{[\xi_2+j\Delta-\frac{\Delta}{2}, \xi_2+j\Delta+\frac{\Delta}{2})}(u_2) h(i\Delta, j\Delta),
\end{aligned}$$

and  $\Delta$  is the grid size while  $p, \tilde{p}$  are the kernel truncation parameters at the field and volatility layers respectively.

Here, we give an analytical formula for the mean squared error (MSE) involved. Since this is difficult to evaluate in practice, we also give an upper bound which is useful in its own right. The corresponding proofs are given in the Appendix.

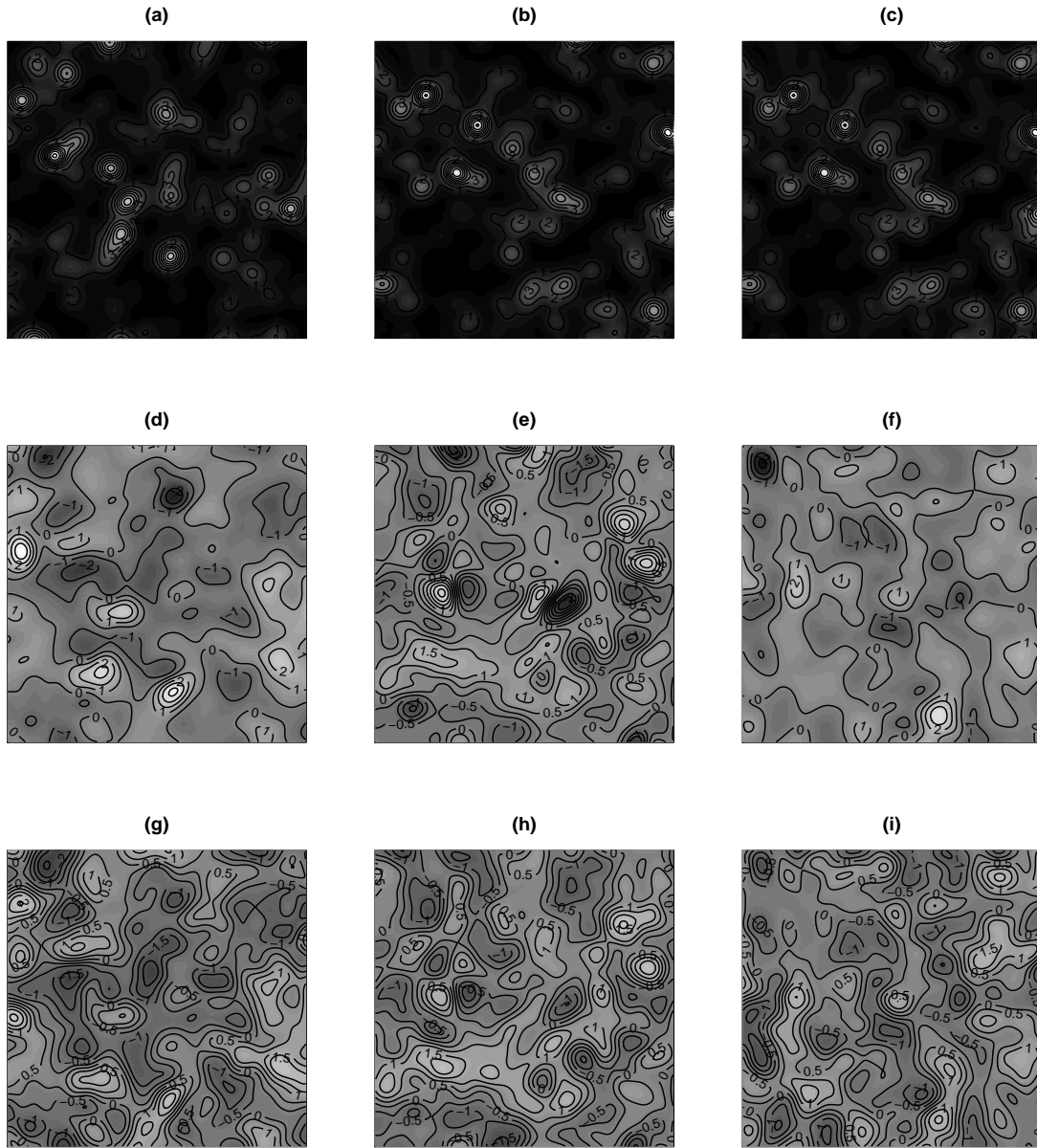


Figure 2: Contour plots for three simulations from Model (6): (a)-(c) the stochastic volatility with  $\eta = 4$ ; (d)-(f) the VMMA with  $\lambda = 4$  and IG basis whose seed has mean  $a = 1$  and variance  $b = 2$ ; (g)-(i) the GMA with the same kernel structure as the VMMA. For the plots in the same column, the VMMA is constructed using associated stochastic volatility and the same realisation of the Gaussian driving noise is used for the VMMA and GMA to facilitate comparison.

**Theorem 3.** Let  $Y$  be a two-tier VMMA where the mean of the Lévy seed  $L'$  is given by  $a > 0$  and let  $Z$  be the TSDC approximation of  $Y$ . Then,  $\mathbb{E}[|Y(\mathbf{x}) - Z(\mathbf{x})|^2] = T1 + T2 + T3$  where:

$$\begin{aligned}
T1 &= a \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} \right) \left( \int_{\mathbb{R}^2} h(\boldsymbol{\xi} - \mathbf{u}) d\mathbf{u} + \sum_{i,j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2 \right) - 2 \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) \mathbb{E}[\sigma(\boldsymbol{\xi}) \sigma_{\Delta}(\boldsymbol{\xi})] d\boldsymbol{\xi}, \\
T2 &= \left[ \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} - \sum_{i,j=-p}^p g^2(i\Delta, j\Delta) \Delta^2 \right) + 2 \sum_{i,j=-p}^p g(i\Delta, j\Delta) \left( g(i\Delta, j\Delta) \Delta^2 - \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right] \\
&\quad \times \left( a \sum_{i',j'=-\tilde{p}}^{\tilde{p}} h(i'\Delta, j'\Delta) \Delta^2 \right), \\
T3 &= 2 \left[ \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) \mathbb{E}[\sigma(\boldsymbol{\xi}) \sigma_{\Delta}(\boldsymbol{\xi})] d\boldsymbol{\xi} - \int_{\mathbb{R}^2} g(\mathbf{x} - \boldsymbol{\xi}) g_{\Delta}(\mathbf{x}, \boldsymbol{\xi}) \mathbb{E}[\sigma(\boldsymbol{\xi}) \sigma_{\Delta}(\boldsymbol{\xi})] d\boldsymbol{\xi} \right. \\
&\quad \left. - a \sum_{i',j'=-\tilde{p}}^{\tilde{p}} h(i'\Delta, j'\Delta) \Delta^2 \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} - \sum_{i,j=-p}^p g(i\Delta, j\Delta) \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right].
\end{aligned}$$

By letting  $\Psi_L(\theta)$  denote the Laplace exponent of  $L'$  evaluated at  $\theta$ , we can also express  $\mathbb{E}[\sigma(\boldsymbol{\xi}) \sigma_{\Delta}(\boldsymbol{\xi})]$  as:

$$\frac{1}{4\pi} \int_0^\infty \int_0^\infty \left[ 1 - e^{\int_{\mathbb{R}^2} \Psi_L(xh(\boldsymbol{\xi}-\mathbf{u})) d\mathbf{u}} - e^{\int_{\mathbb{R}^2} \Psi_L(yh_{\Delta}(\boldsymbol{\xi}, \mathbf{u})) d\mathbf{u}} + e^{\int_{\mathbb{R}^2} \Psi_L(xh(\boldsymbol{\xi}-\mathbf{u}) + yh_{\Delta}(\boldsymbol{\xi}, \mathbf{u})) d\mathbf{u}} \right] \frac{dx dy}{x^{3/2} y^{3/2}}.$$

*Remark 1.*  $T1$  quantifies the part of the MSE that arises from the discrete convolution approximation of  $\sigma(\boldsymbol{\xi})$  since if  $\sigma_{\Delta} \rightarrow \sigma$ , we expect it to decrease to zero. On the other hand,  $T2$  gives us the part of the MSE that can be attributed to the kernel discretisation and truncation of  $g$  since if  $g_{\Delta} \rightarrow g$ ,  $T2$  would decrease to zero. The error from the combined effect of simulating  $\sigma$  and using  $g_{\Delta}$  is represented by  $T3$ .

*Remark 2.* Theorem 3 gives us a semi-explicit formula for the MSE involved in our simulations. Although we can approximate this by numerical integrations, in general, it is hard to obtain a full analytic expression for this due to the Laplace exponent of  $L'$ . For example, if  $L'$  has an IG distribution with mean and shape parameter  $\mu > 0$  and  $\alpha > 0$ :

$$\Psi_L(\theta) = \frac{\alpha}{\mu} \left( 1 - \sqrt{1 + \frac{2\mu^2\theta}{\alpha}} \right) \Rightarrow e^{\int_{\mathbb{R}^2} \Psi_L(xh(\boldsymbol{\xi}-\mathbf{u})) d\mathbf{u}} = \exp \left( \frac{\alpha}{\mu} \int_{\mathbb{R}^2} \left( 1 - \sqrt{1 + \frac{2\mu^2 x h(\boldsymbol{\xi}-\mathbf{u})}{\alpha}} \right) d\mathbf{u} \right),$$

which is hard to simplify in general.

*Remark 3.* If instead of modelling  $\sigma^2$ , we modelled  $\sigma$  or  $\log(\sigma^2)$ ,  $\mathbb{E}[\sigma(\boldsymbol{\xi}) \sigma_{\Delta}(\boldsymbol{\xi})]$  would be easier to evaluate. The disadvantage of the latter strategies is that higher order moments or moments of transformations need to be used for the inference in Section 5.

Although it is hard to calculate the MSE in practice due to the presence of the term  $\mathbb{E}[\sigma(\boldsymbol{\xi}) \sigma_{\Delta}(\boldsymbol{\xi})]$ , we can obtain a useful upper bound by using the relationship between the harmonic, arithmetic and geometric means of  $\sigma^2$  and  $\sigma_{\Delta}^2$ .

**Corollary 3.** Let  $Y$  be a two-tier VMMA where the mean of the Lévy seed  $L'$  is given by  $a > 0$  and let  $Z$  be the TSDC approximation of  $Y$ . Then,  $\mathbb{E}[|Y(\mathbf{x}) - Z(\mathbf{x})|^2] \leq T2 + T4 + T5$  where  $T2$  is as defined in Theorem 3, and:

$$\begin{aligned}
T4 &= \left( a \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} \right) \frac{\left( \int_{\mathbb{R}^2} h(\boldsymbol{\xi} - \mathbf{u}) d\mathbf{u} - \sum_{i=-\tilde{p}}^{\tilde{p}} \sum_{j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2 \right)^2}{\int_{\mathbb{R}^2} h(\boldsymbol{\xi} - \mathbf{u}) d\mathbf{u} + \sum_{i=-\tilde{p}}^{\tilde{p}} \sum_{j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2}, \\
\text{while } T5 &= a \left[ \int_{\mathbb{R}^2} h(\boldsymbol{\xi} - \mathbf{u}) d\mathbf{u} - \sum_{i,j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2 \right] \left[ \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} - \sum_{i,j=-p}^p g^2(i\Delta, j\Delta) \Delta^2 \right) \right]
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i,j=-p}^p g(i\Delta, j\Delta) \left( g(i\Delta, j\Delta) \Delta^2 - \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \\
& + \left( \int_{\mathbb{R}^2} h(\boldsymbol{\xi} - \mathbf{u}) d\mathbf{u} - \sum_{i',j'=-\tilde{p}}^{\tilde{p}} h(i'\Delta, j'\Delta) \Delta^2 \right) \left( \sum_{i,j=-p}^p g(i\Delta, j\Delta) \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \Bigg].
\end{aligned}$$

By using the following assumption, we can derive a result to help us find corresponding orders of convergence:

**Assumption 1.**  $g^2(\mathbf{w})$  and  $h^2(\mathbf{w})$  have bounded Hessian matrices.

**Lemma 1.** Let  $R = p\Delta$  and  $\tilde{R} = \tilde{p}\Delta$  be the fixed truncation ranges for  $g$  and  $h$  respectively. Under Assumption 1:

$$\begin{aligned}
\int_{\mathbb{R}^2} g^2(\mathbf{w}) d\mathbf{w} - \sum_{i,j=-p}^p g^2(i\Delta, j\Delta) \Delta^2 &= O(\Delta^2), \quad g(i\Delta, j\Delta) \Delta^2 - \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} = O(\Delta^4), \\
\text{and } \int_{\mathbb{R}^2} h(\mathbf{w}) d\mathbf{w} - \sum_{i,j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2 &= O(\Delta^2).
\end{aligned}$$

The analysis so far has been for fixed  $R$  and  $\tilde{R}$ . Further suppose that:

**Assumption 2.**  $R = O(\Delta^{-r})$  and  $\tilde{R} = O(\Delta^{-r})$  where  $0 < r < 3$ .

Now,  $R, \tilde{R} \rightarrow \infty$  as  $\Delta \rightarrow 0$  so that the MSE converges to zero. As shown in the next two examples, the order of this convergence will depend on the forms of  $g$  and  $h$ .

*Example 8.* Since:

$$\begin{aligned}
\int_{\mathbb{R}^2} g^2(\mathbf{w}) d\mathbf{w} - \sum_{i,j=-p}^p g^2(i\Delta, j\Delta) \Delta^2 &= \left( \int_{\mathbb{R}^2} g^2(\mathbf{w}) d\mathbf{w} - \int_{-(R+\Delta/2)}^{R+\Delta/2} \int_{-(R+\Delta/2)}^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w} \right) \\
&+ \left( \int_{-(R+\Delta/2)}^{R+\Delta/2} \int_{-(R+\Delta/2)}^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w} - \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} g^2(i\Delta, j\Delta) \Delta^2 \right),
\end{aligned}$$

we find the order of convergence of  $\int_{\mathbb{R}^2} g^2(\mathbf{w}) d\mathbf{w} - \int_{-(R+\Delta/2)}^{R+\Delta/2} \int_{-(R+\Delta/2)}^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w}$  for  $R = O(\Delta^{-r})$  and  $\Delta \rightarrow 0$ . Suppose that  $g(w_1, w_2) = g(-w_1, w_2) = g(w_1, -w_2)$ , i.e.  $g$  is symmetric about the axes, and  $g$  is bounded over  $\mathbb{R}^2$ . In addition, for a large and fixed value of  $|w_1|$ ,  $g(\mathbf{w}) \sim A_1(|w_1|)|w_2|^{-\beta}$  for some square integrable function  $A_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  and  $\beta > 1/2$ . Similarly, for a large and fixed value of  $|w_2|$ ,  $g(\mathbf{w}) \sim A_2(|w_2|)|w_1|^{-\alpha}$  for some square integrable function  $A_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  and  $\alpha > 1/2$ . Notice that this implies that  $g$  behaves proportional to  $|w_1|^{-\alpha}|w_2|^{-\beta}$  for large  $|w_1|$  and  $|w_2|$ . Applying these conditions, we have:

$$\begin{aligned}
& \int_{\mathbb{R}^2} g^2(\mathbf{w}) d\mathbf{w} - \int_{-(R+\Delta/2)}^{R+\Delta/2} \int_{-(R+\Delta/2)}^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w} \\
&= 4 \int_0^\infty \int_{R+\Delta/2}^\infty g^2(\mathbf{w}) d\mathbf{w} + 4 \int_{R+\Delta/2}^\infty \int_0^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w} \text{ by the symmetry about the axes,} \\
&= 4 \left[ \int_{R+\Delta/2}^\infty \int_{R+\Delta/2}^\infty g^2(\mathbf{w}) d\mathbf{w} + \int_0^{R+\Delta/2} \int_{R+\Delta/2}^\infty g^2(\mathbf{w}) d\mathbf{w} + \int_{R+\Delta/2}^\infty \int_0^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w} \right] \\
&\sim 4 \left[ \int_{R+\Delta/2}^\infty \int_{R+\Delta/2}^\infty C_1 |w_1|^{-2\alpha} |w_2|^{-2\beta} d\mathbf{w} + \int_0^{R+\Delta/2} \int_{R+\Delta/2}^\infty A_2^2(|w_2|) |w_1|^{-2\alpha} d\mathbf{w} + \int_{R+\Delta/2}^\infty \int_0^{R+\Delta/2} A_1^2(|w_1|) |w_2|^{-2\beta} d\mathbf{w} \right] \\
&< 4 \left[ \frac{C_1 (R + \Delta/2)^{2-2\alpha-2\beta}}{(2\alpha-1)(2\beta-1)} + \frac{C_2 (R + \Delta/2)^{1-2\alpha}}{2\alpha-1} + \frac{C_3 (R + \Delta/2)^{1-2\beta}}{2\beta-1} \right] \\
&= O(\Delta^{r(2\min(\alpha, \beta)-1)}),
\end{aligned}$$

where  $C_1$  is a finite constant,  $C_2 = \int_0^\infty A_2^2(|w_2|)dw_2$  and  $C_3 = \int_0^\infty A_1^2(|w_1|)dw_1$ . Adding this to the previous bound that we obtained by assuming a bounded Hessian for  $g$ , we have:

$$\int_{\mathbb{R}^2} g^2(\mathbf{w})d\mathbf{w} - \sum_{i,j=-p}^p g^2(i\Delta, j\Delta) \Delta^2 < O(\Delta^{\min(2, r(2\min(\alpha, \beta)-1))}).$$

If  $h$  is symmetric about the axes and shares the same asymptotic properties as  $g$  but with parameters  $\tilde{\alpha}$  and  $\tilde{\beta}$  in place of  $\alpha$  and  $\beta$  respectively, we can use an analogous approach to obtain  $\int_{\mathbb{R}^2} h(\mathbf{w})d\mathbf{w} - \sum_{i,j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2 < O(\Delta^{\min(2, r(2\min(\tilde{\alpha}, \tilde{\beta})-1))})$ .

Applying these bounds to  $T2$ ,  $T4$  and  $T5$ , we have:

$$\begin{aligned} T2 &= O(\Delta^{\min(2, r(2\min(\alpha, \beta)-1), 3-r)}), \quad T4 = O(\Delta^{2\min(2, r(2\min(\tilde{\alpha}, \tilde{\beta})-1)}), \\ \text{and } T5 &= O(\Delta^{\min(4, 2r(2\min(\tilde{\alpha}, \tilde{\beta})-1), 2r(\min(\alpha, \beta)+\min(\tilde{\alpha}, \tilde{\beta})-1), 2+r(2\min(\alpha, \beta)-1), 5-r)}), \\ \Rightarrow \mathbb{E} [|Y(\mathbf{x}) - Z(\mathbf{x})|^2] &\leq O(\Delta^{\min(2, r(2\min(\alpha, \beta)-1), 2r(2\min(\tilde{\alpha}, \tilde{\beta})-1), 3-r)}). \end{aligned}$$

Note that since  $p = \lfloor R/\Delta \rfloor = O(\Delta^{-r-1})$ , in  $T2$  and  $T5$ , we have:

$$\begin{aligned} &\sum_{i,j=-p}^p g(i\Delta, j\Delta) \left( g(i\Delta, j\Delta) \Delta^2 - \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w})d\mathbf{w} \right) \\ &< \sup_{\mathbf{w} \in \mathbb{R}^2} g(\mathbf{w}) \sum_{i,j=-p}^p \left( g(i\Delta, j\Delta) \Delta^2 - \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w})d\mathbf{w} \right) = O(\Delta^{3-r}). \end{aligned}$$

*Example 9.* If we assume that  $g$  and  $h$  are isotropic, and  $g(\mathbf{w}) \sim |\mathbf{w}|^{-\chi}$  while  $h(\mathbf{w}) \sim |\mathbf{w}|^{-\tilde{\chi}}$  for some  $\chi, \tilde{\chi} > 1/2$  when  $|\mathbf{w}|$  is large, a simpler convergence bound for the MSE can be obtained. With  $C$  being a constant:

$$\begin{aligned} &\int_{\mathbb{R}^2} g^2(\mathbf{w})d\mathbf{w} - \int_{-(R+\Delta/2)}^{R+\Delta/2} \int_{-(R+\Delta/2)}^{R+\Delta/2} g^2(\mathbf{w})d\mathbf{w} \leq \int_{|\mathbf{w}| > R+\Delta/2} g^2(|\mathbf{w}|)d|\mathbf{w}| = \frac{C(R+\Delta/2)^{1-2\chi}}{2\chi-1} = O(\Delta^{r(2\chi-1)}). \\ \Rightarrow \int_{\mathbb{R}^2} g^2(\mathbf{w})d\mathbf{w} - \sum_{i,j=-p}^p g^2(i\Delta, j\Delta) \Delta^2 &\leq O(\Delta^{\min(2, r(2\chi-1))}). \end{aligned}$$

Similarly, one can show that  $\int_{\mathbb{R}^2} h(\mathbf{w})d\mathbf{w} - \sum_{i,j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2 = O(\Delta^{\min(2, r(2\tilde{\chi}-1))})$ .

Correspondingly, we have:

$$\begin{aligned} T2 &= O(\Delta^{\min(2, r(2\chi-1), 3-r)}), \quad T4 = O(\Delta^{2\min(2, r(2\tilde{\chi}-1))}) \text{ and } T5 = O(\Delta^{\min(4, 2r(2\tilde{\chi}-1), 2r(\chi+\tilde{\chi}-1), 2+r(2\chi-1), 5-r)}), \\ \Rightarrow \mathbb{E} [|Y(\mathbf{x}) - Z(\mathbf{x})|^2] &\leq O(\Delta^{\min(2, r(2\chi-1), 2r(2\tilde{\chi}-1), 3-r)}). \end{aligned}$$

*Example 10.* For illustration, we provide the explicit MSE upper bound for Model (6). With  $\Phi(x; y)$  denoting the standard Normal distribution function evaluated between  $\mathbf{x}$  and  $\mathbf{y}$ , the contributing terms are:

$$\begin{aligned} T4 &= \frac{a\lambda}{2\pi} \left[ 1 + \frac{\eta}{\pi} \left[ 1 + 2 \sum_{i'=1}^{\tilde{p}} \exp(-\eta\Delta^2 i'^2) \right]^2 \Delta^2 - 4 \left( 1 + \frac{\pi}{\eta \left[ 1 + 2 \sum_{i'=1}^{\tilde{p}} \exp(-\eta\Delta^2 i'^2) \right]^2 \Delta^2} \right)^{-1} \right] \\ T2 &= \frac{a\lambda\eta}{\pi^2} \Delta^2 \left[ 1 + 2 \sum_{i'=1}^{\tilde{p}} \exp(-\eta\Delta^2 i'^2) \right]^2 \\ &\quad \times \left[ \frac{1}{2} + \frac{\lambda}{\pi} \Delta^2 \left[ 1 + 2 \sum_{i=1}^p \exp(-2\lambda\Delta^2 i^2) \right]^2 - 2 \left[ \sum_{i=-p}^p e^{-\lambda\Delta^2 i^2} \Phi \left( \sqrt{2\lambda} \left( i\Delta - \frac{\Delta}{2} \right), \sqrt{2\lambda} \left( i\Delta + \frac{\Delta}{2} \right) \right) \right]^2 \right], \end{aligned}$$

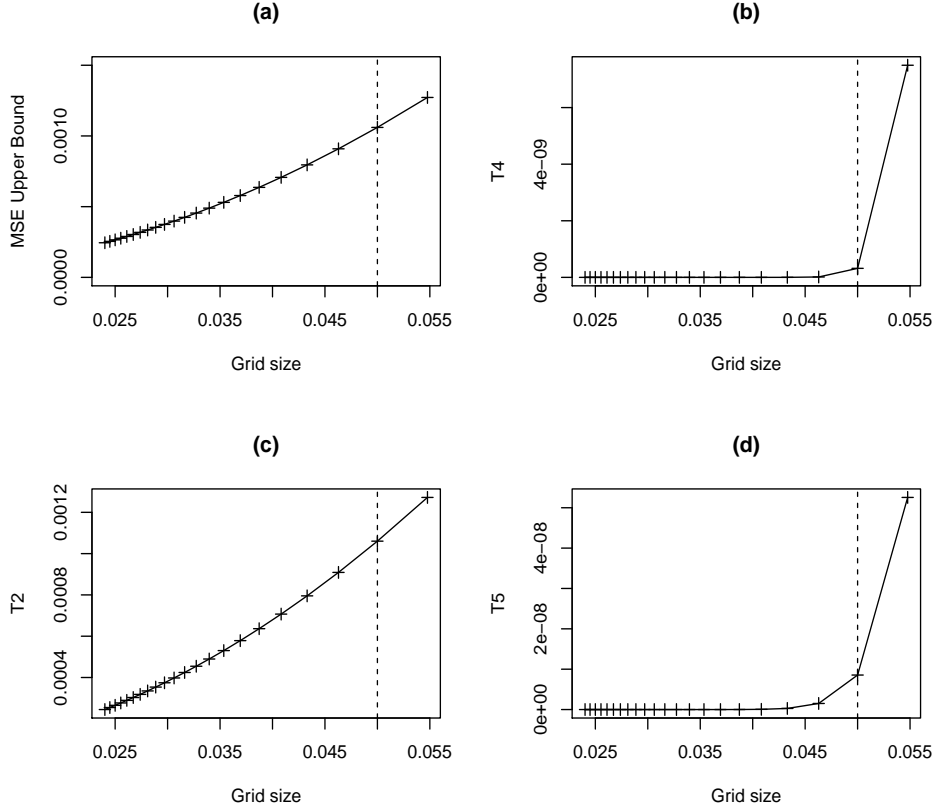


Figure 3: (a) MSE upper bound against  $\Delta$  for Model (6) when we set  $\lambda = \eta = 4$ ,  $a = 1$ ,  $b = 2$ ,  $p = \tilde{p}$  and  $R = K\Delta^{-1}$  where  $K = 0.05^2 \times 30$ ; and the corresponding values of (b)  $T4$ , (c)  $T2$  and (d)  $T5$ . The vertical dotted lines mark our chosen simulation setting.

$$\text{and } T5 = \frac{a\lambda}{\pi} \left[ \frac{1}{2} - \frac{\eta}{2\pi} \left[ 1 + 2 \sum_{i'=1}^{\tilde{p}} \exp(-\eta\Delta^2 i'^2) \right] \Delta^2 - 4 \left[ \sum_{i=-p}^p e^{-\lambda\Delta^2 i^2} \Phi \left( \sqrt{2\lambda} \left( i\Delta - \frac{\Delta}{2} \right), \sqrt{2\lambda} \left( i\Delta + \frac{\Delta}{2} \right) \right) \right]^2 \right. \\ \left. \times \left( \left( 1 + \frac{\pi}{\eta \left[ 1 + 2 \sum_{i'=1}^{\tilde{p}} \exp(-\eta\Delta^2 i'^2) \right]^2 \Delta^2} \right)^{-1} - \frac{\eta}{2\pi} \Delta^2 \left[ 1 + 2 \sum_{i'=1}^{\tilde{p}} \exp(-\eta\Delta^2 i'^2) \right]^2 \right) \right].$$

Using our simulation settings, i.e.  $\lambda = \eta = 4$ ,  $a = 1$ ,  $b = 2$  and  $p = \tilde{p}$ , we examine the behaviour of  $T2$ ,  $T4$  and  $T5$ , as well as the resulting upper bound on the MSE for the case:  $R = K\Delta^{-1}$  where  $K = 0.05^2 \times 30$  so that as  $\Delta$  decreases,  $R$  increases. We know from Examples 8 and 9 that the MSE converges to zero at most as fast as  $O(\Delta^2)$  when  $R = O(\Delta^{-1})$  since the squared exponential dominates any power function.

Figure 3 shows the upper bound,  $T4$ ,  $T2$  and  $T5$  values. We notice that  $T2$  is much larger in magnitude than  $T4$  and  $T5$ . This indicates that for our choice of  $a = 1$  and  $\lambda = \eta = 4$ , the error due to the kernel truncation and discretisation of  $g$  outweighs those due to the simulation of  $\sigma$ . As such, the asymptotic behaviour of our upper bound is driven largely by the behaviour of  $T2$ . The vertical dotted lines in each plot indicate our simulation choice of  $\Delta = 0.05$  and  $R = 1.5$  for the experiments in Section 5. From the plots, we also see that when  $R$  increases and  $\Delta$  decreases,  $T4$ ,  $T2$  and  $T5$  converge smoothly to 0 so that the MSE upper bound (and the MSE itself) converges to zero as expected.

*Remark 4.* So far, we have assumed that  $g$  takes finite values over  $\mathbb{R}^2$ . This holds for many choices of  $g$ . In the case of a singularity at zero which occurs for the Matérn kernel (3) with  $(\alpha - d)/2 \in (-1/2, 0)$ , a so-called hybrid scheme similar to that in Bennedsen et al. (2015) can be used.

## 5 Inference

Before we introduce the two-step moments-matching method for VMMA, we will look at several classical inference approaches and give reasons why it is difficult to implement them for these models.

### 5.1 A note on classical methods

In Econometrics where we have high-frequency financial data, a tool frequently used to estimate the parameters of the volatility is that of realised variance. This involves the sum of the squared increments of the log-price over small time intervals. Recently, this concept has been extended to two-dimensions in Pakkanen (2014). Now instead of taking increments over time intervals, we take increments over rectangles in our data region. Under certain assumptions, one can show that this two-dimensional realised variance properly normalised, converges to a weighted integral of the volatility as these rectangles get smaller. This could enable us to estimate the parameters of the volatility by moments-based or quasi-likelihood methods. Unfortunately, it was found that not all the required conditions hold for general VMMA and it is hard to establish the convergence.

Another classical approach would be to use likelihood-based or Bayesian inference. In general, however, since we do not know the specific distribution of  $\sigma^2$  and an approximation of the conditional variance does not guarantee valid covariance matrices, implementing such strategies for a VMMA is not straightforward.

Alternatively, one might consider a direct moments-matching method involving higher order moments. As we have found though, this may lead to parameter sign errors. That is, we could estimate a positive parameter as negative if no adjustments are made.

Since these classical approaches are hard to implement, we will develop a two-step moments-matching estimation method for our VMMA.

### 5.2 Moments-based estimation

To illustrate our moments-based method, we use data generated from Model (6). Recall that our VMMA  $Y$  is stationary, but when we condition on  $\sigma^2$ ,  $Y$  is non-stationary. Specifically, when we express its observations as a vector,  $Y|\sigma^2 \sim N(0, V)$  where  $V = (V_{ij})$  with  $V_{ij} = \int_{\mathbb{R}^2} \lambda^2 \pi^{-2} \exp\left(-\lambda(\mathbf{x}_i - \boldsymbol{\xi})^T(\mathbf{x}_i - \boldsymbol{\xi}) - \lambda(\mathbf{x}_j - \boldsymbol{\xi})^T(\mathbf{x}_j - \boldsymbol{\xi})\right) \sigma^2(\boldsymbol{\xi}) d\boldsymbol{\xi}$  and where  $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^2$  are two data locations. Since  $\sigma^2$  is stationary, we should expect to observe stationarity in  $Y$  over a large region. This should allow us to estimate  $\lambda$  and  $a$  through the empirical normalised variogram and variance.

By definition, the normalised variogram of our VMMA,  $Y(\mathbf{x})$ , is given by:

$$\gamma(d_{\mathbf{x}}) := \frac{\mathbb{E}[(Y(\mathbf{x}) - Y(\mathbf{x}^*))^2]}{\text{Var}(Y(\mathbf{x}))} = 2(1 - \text{Corr}(Y(\mathbf{x}), Y(\mathbf{x}^*))) = 2\left(1 - \exp\left(-\frac{\lambda d_{\mathbf{x}}^2}{2}\right)\right), \quad (14)$$

where  $\mathbf{x}, \mathbf{x}^* \in \mathbb{R}^2$  and  $|\mathbf{x} - \mathbf{x}^*| = d_{\mathbf{x}}$ .

Let  $(Y(x_1^i, x_2^j)) \in \mathbb{R}^{M \times M}$  be our data matrix and  $Y(\mathbf{x}_k) \in \mathbb{R}^{M^2}$  be its corresponding data vector. With  $N(d_{\mathbf{x}})$  denoting the set containing all the pairs of indices of sites with spatial distance  $d_{\mathbf{x}}$ , we can estimate the normalised variogram by:

$$\hat{\gamma}(d_{\mathbf{x}}) = \frac{1}{|N(d_{\mathbf{x}})|} \sum_{(k,l) \in N(d_{\mathbf{x}})} \frac{(Y(\mathbf{x}_k) - Y(\mathbf{x}_l))^2}{\hat{\kappa}_2}, \quad (15)$$

where  $\hat{\kappa}_2 = \frac{1}{M^2-1} \sum_{k=1}^{M^2} (Y(\mathbf{x}_k) - \bar{Y})^2$  and  $\bar{Y} = \frac{1}{M^2} \sum_{k=1}^{M^2} Y(\mathbf{x}_k)$ . By matching (14) and (15), we can estimate the rate parameter of our field by:

$$\hat{\lambda} = -\frac{2 \log(1 - \hat{\gamma}(\Delta)/2)}{\Delta^2},$$



where  $\Delta$  is the simulation grid size. For Model (6),  $\kappa_2 = \frac{a\lambda}{2\pi}$ . Thus, with  $\hat{\lambda}$  at hand, we can estimate  $a$  by:

$$\hat{a} = \frac{2\pi\hat{\kappa}_2}{\hat{\lambda}}.$$

Next, we want to obtain estimates for the parameters  $b$  and  $\eta$ . These determine the extent of non-stationarity in  $Y|\sigma^2$  which is shown through differing variance and covariance structures across subregions. Thus, it seems natural to infer about  $b$  and  $\eta$  by comparing estimated local variances. This requires some sort of subsetting. To retain the correlation between our local variance estimates for the next step of our inference, we use a moving window strategy.

Figure 4 illustrates how the  $q \times q$  moving window, which is represented by the small box on the bottom left corner of Plot (a), selects data points for local variance calculations to form a field of estimates in Plot (b). The current location at which the local variance is being calculated is represented by the circle and labelled  $\xi_1$ . By moving the window from left to right and then up the rows of the  $M \times M$  data matrix, we obtain the  $(M - q + 1) \times (M - q + 1)$  field of local variance estimates, as illustrated in Plot (b). The parameter  $q$  is a tuning parameter in our inference method. As will be evident later, different  $q$  values lead to different inferred volatility cluster sizes which in turn are related to the values of  $b$  and  $\eta$ . For  $\mathbf{x} \in \mathbb{R}^2$ , we can index the local variance estimator by  $Q = (q - 1)/2$  as follows:

$$\hat{\sigma}_I^2(\mathbf{x}, Q) = \frac{1}{(2Q + 1)^2} \sum_{l=-Q}^Q \sum_{k=-Q}^Q Y^2(\mathbf{x} + (l, k)\Delta).$$

The local variances estimates are estimates for the conditional variance at the centres of the subregions marked out by the moving window procedure. The analytical formula for the latter is given by:

$$\sigma_I^2(\xi_i) = \int_{\mathbb{R}^2} \frac{\lambda^2}{\pi^2} \exp\left(-2\lambda(\xi_i - \xi)^T(\xi_i - \xi)\right) \sigma^2(\xi) d\xi, \quad (16)$$

where  $\xi_i$  denotes the centre of the subregion  $A_i$  for  $i = 1, \dots, \widetilde{M}$  and  $\widetilde{M} = (M - q + 1)^2$ .

By comparing the mean of our local variances to  $a\lambda(2\pi)^{-1}$  and writing  $\bar{\sigma}_I^2 = \frac{1}{\widetilde{M}} \sum_{i=1}^{\widetilde{M}} \hat{\sigma}_I^2(\xi_i, Q)$ , we get another estimator for  $a$ :

$$\hat{a}_2 = \frac{2\pi\bar{\sigma}_I^2}{\hat{\lambda}}.$$

From the proof of Example 7,  $\text{Cov}(\sigma_I^2(\xi_i), \sigma_I^2(\xi_j)) = A \exp(-B(\xi_i - \xi_j)^T(\xi_i - \xi_j))$ , where  $A = b\lambda^3\eta(4\pi^3(2\lambda + \eta))^{-1}$  and  $B = \lambda\eta(2\lambda + \eta)^{-1}$ . Using the empirical variance and normalised variogram at the first lag of  $\hat{\sigma}_I^2, \hat{\psi}(\Delta)$ , we obtain:

$$\hat{A} = \frac{1}{\widetilde{M} - 1} \sum_{i=1}^{\widetilde{M}} (\hat{\sigma}_I^2(\xi_i) - \bar{\sigma}_I^2)^2 \text{ and } \hat{B} = -\frac{\log(1 - \hat{\psi}(\Delta)/2)}{\Delta^2}.$$

This in turn gives us:

$$\hat{b} = \frac{4\pi^3\hat{A}}{\hat{\lambda}^2\hat{B}} \text{ and } \hat{\eta} = \frac{2\hat{\lambda}\hat{B}}{\hat{\lambda} - \hat{B}}.$$

*Remark 5.* We have used Model (6) to illustrate our method. More generally, this strategy works when we have parameters representing the variance and correlation of  $Y$  and  $\sigma^2$  respectively, and when an analytical expression for  $\text{Cov}(\sigma_I^2(\xi_i), \sigma_I^2(\xi_j))$  is available.

### 5.3 Theoretical properties of the estimators

In this subsection, we derive some properties of the local variance and moments-based parameter estimators. We show that under suitable conditions, the latter are consistent. Proofs of the results, if not shown, can be found in the Appendix.

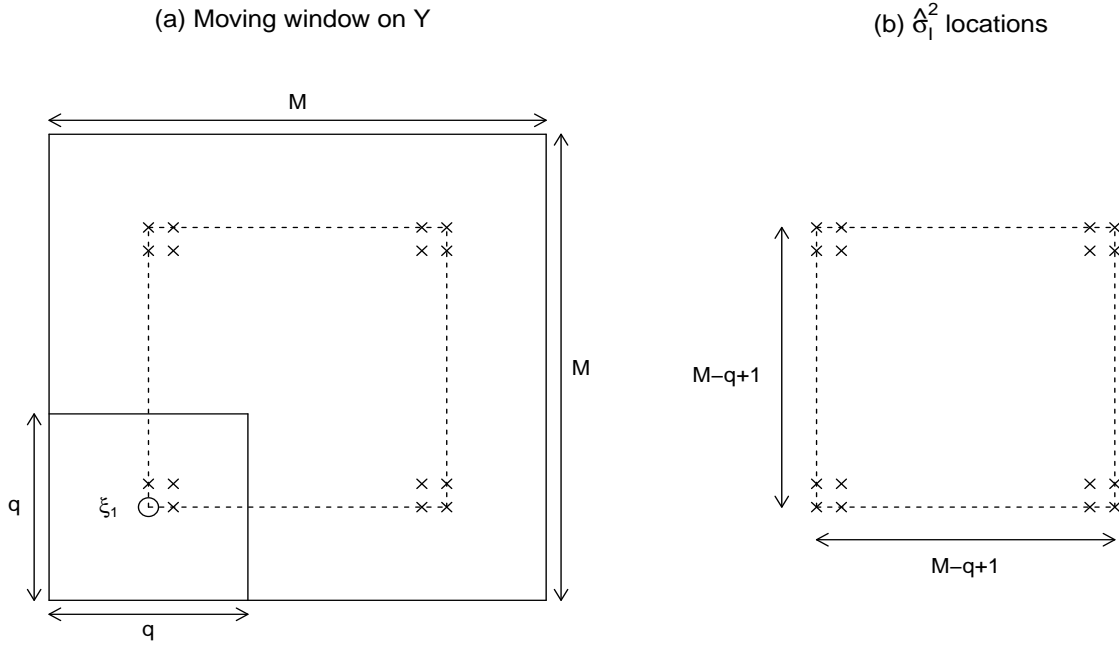


Figure 4: Moving window strategy: (a) The  $q \times q$  moving window, represented by the smaller box on the bottom left corner, slides across the  $M \times M$  data region to select the data points for local variance estimation. (b) A  $(M - q + 1) \times (M - q + 1)$  field of local variance estimates ( $\hat{\sigma}_I^2$ ) is created.

**Theorem 4.** Suppose that the following conditions hold:

1.

$$\frac{\sum_{l',k',l,k=-Q}^Q \mathbb{E} \left[ \left( \int_{\mathbb{R}^2} g(\mathbf{x} + (l, k)\Delta - \xi) g(\mathbf{x} + (l', k')\Delta - \xi) \sigma^2(\xi) d\xi \right)^2 \right]}{(2Q + 1)^4} \rightarrow 0;$$

2.  $C(d_{x_1}, d_{x_2}) := \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (d_{x_1}, d_{x_2})))$  has a finite gradient  $\nabla C = (\partial C / \partial d_{x_1}, \partial C / \partial d_{x_2})$  over  $\mathbb{R}^2$

Then, the local variance estimator  $\hat{\sigma}_I^2(\mathbf{x}, Q) \xrightarrow{\mathcal{L}_2} \sigma_I^2(\mathbf{x})$  when  $Q \rightarrow \infty$  and  $\Delta = O(Q^{-\tilde{r}})$  for  $\tilde{r} > 0$ .

**Remark 6.** The conditions “ $Q \rightarrow \infty$  and  $\Delta = O(Q^{-\tilde{r}})$  for  $\tilde{r} > 0$ ” means that we require infill asymptotics. If  $0 < \tilde{r} < 1$ , the range of our moving window,  $\tau = Q\Delta$  increases as  $Q \rightarrow \infty$ . On the other hand, if  $\tilde{r} = 1$ , we have a fixed range and if  $\tilde{r} > 1$ ,  $\tau \rightarrow 0$  as  $Q \rightarrow \infty$ . Based on the proof of Theorem 4, the rate of the  $\mathcal{L}_2$  convergence of our estimator increases as  $\tilde{r}$  increases.

**Corollary 4.** When Theorem 4 holds, the mean, variance and normalised variogram of the estimated local variance field converge to those of the true conditional variance,  $\sigma_I^2$ .

**Example 11.** We show that the assumptions required for Theorem 4 hold for Model (6). From Example 7, we know that:

$$\text{Cov}(Y^2(\mathbf{x}), Y^2(\mathbf{x}^*)) = \frac{(b\lambda\eta + a^2(2\lambda + \eta)\pi)\lambda^2}{2\pi^3(2\lambda + \eta)} \exp\left(-\lambda(\mathbf{x} - \mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*)\right) + A \exp\left(-B(\mathbf{x} - \mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*)\right).$$

This implies that (21) in the Appendix is equal to:

$$\begin{aligned} & \frac{1}{(2Q + 1)^4} \sum_{l',k',l,k=-Q}^Q [\text{Cov}(Y^2(\mathbf{x} + (l, k)\Delta), Y^2(\mathbf{x} + (l', k')\Delta)) - \text{Cov}(\sigma_I^2(\mathbf{x} + (l, k)\Delta), \sigma_I^2(\mathbf{x} + (l', k')\Delta))] \\ &= \frac{1}{(2Q + 1)^4} \sum_{l',k',l,k=-Q}^Q \frac{(b\lambda\eta + a^2(2\lambda + \eta)\pi)\lambda^2}{2\pi^3(2\lambda + \eta)} \exp\left(-\lambda\Delta^2[(l - l')^2 + (k - k')^2]\right) \end{aligned}$$

$$\begin{aligned}
&= \frac{(b\lambda\eta + a^2(2\lambda + \eta)\pi)\lambda^2}{2\pi^3(2\lambda + \eta)(2Q + 1)^4} \left[ \sum_{k=-Q}^Q \sum_{k'=-Q}^Q \exp\left(-\lambda\Delta^2(k - k')^2\right) \right]^2 \\
&= \frac{(b\lambda\eta + a^2(2\lambda + \eta)\pi)\lambda^2}{2\pi^3(2\lambda + \eta)(2Q + 1)^4} \left[ (2Q + 1) + \sum_{\substack{\{k, k' \in \{-Q, Q\}: \\ |k - k'| = 1\}}} e^{-\lambda\Delta^2(k - k')^2} + \dots + \sum_{\substack{\{k, k' \in \{-Q, Q\}: \\ |k - k'| = 2Q\}}} e^{-\lambda\Delta^2(k - k')^2} \right]^2 \\
&= \frac{(b\lambda\eta + a^2(2\lambda + \eta)\pi)\lambda^2}{2\pi^3(2\lambda + \eta)(2Q + 1)^4} \left[ (2Q + 1) + 4Q \exp(-\lambda\Delta^2) + \dots + 2 \exp(-\lambda\Delta^2(2Q)^2) \right]^2.
\end{aligned}$$

All of the terms in the square brackets of the last line behave like  $O(Q)$  when  $Q$  tends to infinity and  $\Delta$  behaves like  $O(Q^{-\tilde{r}})$  for  $\tilde{r} > 0$ . So, term (21) behaves like  $O(Q^{-2})$  and converges to zero as required.

Next, we show that the second condition of Theorem 4 holds:

$$C(d_{x_1}, d_{x_2}) = \frac{b\lambda^3\eta}{4\pi^3(2\lambda + \eta)} \exp\left(\frac{-\lambda\eta}{2\lambda + \eta}(d_{x_1}^2 + d_{x_2}^2)\right) \Rightarrow \frac{\partial C}{\partial d_{x_i}} = \frac{-2b\lambda^4\eta^2 d_{x_i}}{4\pi^3(2\lambda + \eta)^2} \exp\left(\frac{-\lambda\eta}{2\lambda + \eta}(d_{x_1}^2 + d_{x_2}^2)\right),$$

for  $i = 1, 2$ . Using L' Hôpital's Rule,  $\partial C / \partial d_{x_i} \rightarrow 0$  as  $d_{x_i} \rightarrow \infty$ . Thus,  $\nabla C = (\partial C / \partial d_{x_1}, \partial C / \partial d_{x_2})$  is finite over  $\mathbb{R}^2$ .

**Theorem 5.** Suppose that Theorem 4 holds, i.e.  $\hat{\sigma}_I^2(\mathbf{x}, Q) \xrightarrow{\mathcal{L}_3} \sigma_I^2(\mathbf{x})$  when  $Q \rightarrow \infty$  and  $\Delta = O(Q^{-\tilde{r}})$  for  $\tilde{r} > 0$ , and that the number of local variance locations,  $\tilde{M} = O(\Delta^{\tilde{\tau}})$  where  $\tilde{\tau} > 0$ . This means that we have both infill and increasing domain asymptotics. Then if Slutsky's conditions for mean and covariance ergodicity hold for  $Y$  and  $\sigma_I^2$ , i.e.:

$$\begin{aligned}
&\lim_{X \rightarrow \infty} \frac{1}{X^2} \int_0^X \int_0^X \text{Cov}(Y(\mathbf{x}), Y(\mathbf{x} + \mathbf{h})) d\mathbf{h} = 0, \quad \lim_{X \rightarrow \infty} \frac{1}{X^2} \int_0^X \int_0^X \text{Cov}^2(Y(\mathbf{x}), Y(\mathbf{x} + \mathbf{h})) d\mathbf{h} = 0 \\
&\lim_{X \rightarrow \infty} \frac{1}{X^2} \int_0^X \int_0^X \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + \mathbf{h})) d\mathbf{h} = 0 \text{ and } \lim_{X \rightarrow \infty} \frac{1}{X^2} \int_0^X \int_0^X \text{Cov}^2(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + \mathbf{h})) d\mathbf{h} = 0,
\end{aligned}$$

the empirical variance and normalised variogram of  $Y$  are consistent. Furthermore, the empirical mean, variance and normalised variogram of  $\hat{\sigma}_I^2$  converge in probability to the respective theoretical quantities of  $\sigma_I^2$ .

*Proof.* As suggested on page 57 of Cressie (1993), under infill and increasing domain asymptotics, Slutsky's conditions for the mean and covariance ergodicity of the discrete process formed by sampling a continuous process converges to those of the continuous process itself. For example:

$$\begin{aligned}
&\lim_{X, N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \text{Cov}\left(Y(\mathbf{x}), Y\left(\mathbf{x} + \left(\frac{iX}{N}, \frac{jX}{N}\right)\right)\right) = 0 \text{ where } X = O(N^{\tilde{s}}) \text{ with } \tilde{s} < 1, \\
&\Leftrightarrow \lim_{X \rightarrow \infty} \frac{1}{X^2} \lim_{N \rightarrow \infty} \frac{X^2}{N^2} \sum_{i=1}^N \sum_{j=1}^N \text{Cov}\left(Y(\mathbf{x}), Y\left(\mathbf{x} + \left(\frac{iX}{N}, \frac{jX}{N}\right)\right)\right) = 0 \\
&\Leftrightarrow \lim_{X \rightarrow \infty} \frac{1}{X^2} \int_0^X \int_0^X \text{Cov}(Y(\mathbf{x}), Y(\mathbf{x} + \mathbf{h})) d\mathbf{h} = 0
\end{aligned}$$

since in the second line,  $X$  is fixed when we vary  $N$ .

Thus, if Slutsky's conditions hold, the sample mean and covariances for  $Y$  and  $\sigma_I^2$  converge to their theoretical values. Since the normalised variograms are formed from the covariances, the empirical variograms are also consistent.

Now we show that the empirical mean, variance and normalised variogram of  $\hat{\sigma}_I^2$  converge to the respective theoretical equivalents of  $\sigma_I^2$ . Under the conditions of Theorem 4,  $\hat{\sigma}_I^2(\mathbf{x}, Q) \xrightarrow{P} \sigma_I^2(\mathbf{x})$  for arbitrary  $\mathbf{x} \in \mathbb{R}^2$ . Since for  $X_n \xrightarrow{P} X$  and  $Y_n \xrightarrow{P} Y$ , we have that  $aX_n + bY_n \xrightarrow{P} aX + bY$ , it follows that the sample mean of an estimated local variance surface at  $\tilde{M}$  locations,  $\tilde{M}^{-1} \sum_{i=1}^{\tilde{M}} \hat{\sigma}_I^2(\mathbf{s}_i) \xrightarrow{P} \tilde{M}^{-1} \sum_{i=1}^{\tilde{M}} \sigma_I^2(\mathbf{s}_i)$  as  $Q \rightarrow \infty$ . Since we have mean ergodicity, this in turn converges in probability to  $\mathbb{E}[\sigma^2(\mathbf{x})]$  as  $\tilde{M} \rightarrow \infty$ . Following similar arguments with the appropriate use of the Continuous Mapping Theorem, we can show that the sample variance and normalised variogram of  $\hat{\sigma}_I^2$  converge in probability to the theoretical

variance and normalised variogram of  $\sigma_I^2$ . □

*Remark 7.* Recall that  $\text{Cov}(Y(\mathbf{x}), Y(\mathbf{x} + \mathbf{h})) = \mathbb{E}[\sigma^2(\mathbf{0})] \int_{\mathbb{R}^2} g(\mathbf{w})g(\mathbf{w} + \mathbf{h})d\mathbf{w}$  and:

$$\begin{aligned} \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + \mathbf{h})) &= \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} g^2(\mathbf{x} - \boldsymbol{\xi})g^2(\mathbf{x} + \mathbf{h} - \boldsymbol{\xi}^*) \text{Cov}(\sigma^2(\boldsymbol{\xi}), \sigma^2(\boldsymbol{\xi}^*)) d\boldsymbol{\xi}d\boldsymbol{\xi}^* \\ &= \int_{\mathbb{R}^2} \left[ \int_{\mathbb{R}^2} g^2(\mathbf{u} - \mathbf{w})g^2(\mathbf{u} + \mathbf{h})d\mathbf{u} \right] \text{Cov}(\sigma^2(\mathbf{w}), \sigma^2(\mathbf{0})) d\mathbf{w}, \end{aligned}$$

where  $\mathbf{w} = \boldsymbol{\xi} - \boldsymbol{\xi}^*$ ,  $\mathbf{u} = \mathbf{x} - \boldsymbol{\xi}^*$  and we have written the covariance of  $\sigma^2$  in terms of the lag  $\mathbf{w}$ . Using these expressions, we can express Slutsky's conditions in terms of  $g$  and the first two moments of  $\sigma^2$ .

**Lemma 2.** *For Model (6), our two-step moments-matching estimators,  $\hat{\lambda}$ ,  $\hat{a}$ ,  $\hat{a}_2$ ,  $\hat{b}$  and  $\hat{\eta}$  are consistent under infill and increasing domain asymptotics.*

*Proof.* Slutsky's ergodic conditions hold for Model (6) since the covariances of  $Y$  and  $\sigma_I^2$  can be written in terms of Gaussian densities. From Example 11, we also know that Theorem 4 holds. So, the empirical variance and normalised variogram of  $Y$  are consistent, and the empirical mean, variance and normalised variogram of  $\hat{\sigma}_I^2$  converge in probability to the theoretical quantities of  $\sigma_I^2$ . By repeated use of Slutsky's Theorem and the Continuous Mapping Theorem, it is easy to show that our parameter estimators are consistent. □

*Remark 8.* We proved the consistency of our parameter estimators for Model (6). As hinted at in Remark 5, these consistency conditions can be checked for the two-step moments-matching estimators of other VMMA when they have parameters representative of the variance and correlation of  $Y$  and those of  $\sigma^2$ , and when an analytical expression for  $\text{Cov}(\sigma_I^2(\boldsymbol{\xi}_i), \sigma_I^2(\boldsymbol{\xi}_j))$  is available.

## 5.4 Estimation in practice

Under infill and increasing domain asymptotics, our parameter estimators are consistent when the range of our moving window behaves in an appropriate manner with respect to the grid size. To use the moments-based method in practice, we need to fix this range by choosing our tuning parameter  $q$ . In this subsection, we introduce a way to select  $q$  by the so-called maximum regional variance, examine the potential and limitations of such a method as well as illustrate why we chose to define the local variance estimator by the mean of the squared data values instead of a sample variance formula.

### 5.4.1 Choosing $q$ via the maximum regional variance

The size of the volatility clusters, i.e. the regions of high volatility, gives us information about the variance and correlation parameters of  $\sigma^2$ . In the case of Model (4), these are  $b$  and  $\eta$  respectively. Thus, we choose  $q$  to identify the volatility cluster size. Let us define the empirical regional variance at  $\boldsymbol{\xi}_i \in \mathbb{R}^2$  by:

$$\hat{\kappa}_2^I(\boldsymbol{\xi}_i) = \frac{1}{|Y_{A_i}| - 1} \sum_{\mathbf{x} \in A_i} (Y(\mathbf{x}) - \bar{Y}_{A_i})^2,$$

where  $Y_{A_i}$  denotes the data values within the subregion  $A_i$  whose centre is  $\boldsymbol{\xi}_i$  and  $\bar{Y}_{A_i}$ , the empirical mean of  $Y_{A_i}$ .

If  $q$  is too small, the similarity of the  $Y$ s within the capture windows will cause the  $\hat{\kappa}_2^I$  values to be small. As we increase  $q$  to the radius of a volatility cluster, there will be capture windows for which the  $Y$ s within have increasingly different values. This causes the maximum regional variance (MRV), i.e. the maximum of the  $\hat{\kappa}_2^I$  values over the  $\boldsymbol{\xi}_i$  values, to increase. If we increase  $q$  further, the unconditional stationarity of our VMMA and the stationarity of the  $\sigma^2$  layer will cause the  $\hat{\kappa}_2^I$  values and hence its maximum over the subregions to drop. Thus, we can identify appropriate values of  $q$  by calculating the MRV for a range of values and choosing the  $q$  values at the MRV peaks.

Figure 5 shows the MRV values of a data set, Data set 1, simulated from Model (6) with  $\lambda = \eta = 4$ ,  $a = 1$  and  $b = 2$  in

Plot (a) and that simulated from a GMA with the same correlation structure and underlying Gaussian noise in Plot (b) for  $q$  ranging from 9 to 51. To generate each data set, we chose  $\Delta = 0.05$  and  $M = 201$ . We see that although both plots feature peaks in MRV, the peak for the VMMA is larger in magnitude. This indicates the presence of stochastic volatility: the higher the magnitude of the peak, the more prominent the associated volatility cluster. On the other hand, while the peak occurs at  $q = 21$  for the VMMA, it occurs at  $q = 33$  for the GMA. The “volatility clusters” identified in the GMA data have a larger radius. Since larger volatility clusters are less distinct over the same data region, this is indicative of its constant volatility. From this observation, in cases of multiple peaks, we prioritise peaks at lower  $q$  values.

#### 5.4.2 Choice of the local variance estimator

Here, we give some reasoning behind our choice of  $\hat{\sigma}_I^2$  as the local variance estimator instead of  $\hat{\kappa}_2^I$ .

Although computing  $\hat{\kappa}_2^I$  allows us to identify the volatility cluster sizes, these estimates do not work well as local variance estimates. In particular, they underestimate the local variances near the centres of volatility clusters. The key difference between  $\hat{\kappa}_2^I$  and  $\hat{\sigma}_I^2(\xi_i)$  is that instead of using  $\bar{Y}_{A_i}$  to estimate regional mean, we set it to the theoretical zero.

Figures 6(b) and (e) show the fields of  $\hat{\kappa}_2^I$  and  $\hat{\sigma}_I^2$  obtained from Data set 1 for  $q = 21$ . Comparing Plots (b) and (e) (and similar plots for other  $q$  values), we notice that while high  $\hat{\kappa}_2^I$  values mean that we are at the boundaries of the volatility clusters, high  $\hat{\sigma}_I^2$  values mean that we are in the clusters themselves. Using the MRV to choose  $q$  enables us to choose a capture window size so that we find the size of the highest volatility cluster which lies away from the boundaries (so as to be captured by the capture window). From the location of the red diamond in Plot (e), we see that the point whose regional variance is the MRV for  $q = 21$  lies at the slope of this cluster. Note that the white border in the heat plots denote areas for which no estimates are obtained from the moving window approach and this widens as  $q$  increases.

#### 5.4.3 Inference results and discussion

We apply our inference procedure to 100 simulated data sets for Model (6). These were generated with random seeds 1 to 100. Figure 7 shows the estimates and selected  $q$  values from the 100 data sets. The outliers are labelled by their data set indices. Since the medians of the estimates (denoted by the bold black horizontal lines) lie close to the true parameter values (represented by the red horizontal lines) in all cases, our inference method works reasonably well.

Despite the promising results, we note that the moving window approach is sensitive to the most prominent estimated volatility cluster which is in turn influenced by the realisation of the underlying Gaussian noise. In individual cases, special care is also required when we have overlapping clusters and sudden surges in amplitude.

In Figure 8(a), we show the true conditional variance surface of Data set 1. Plots (b) and (c) show its estimated local variance surface as well as that for its corresponding GMA data set. For the VMMA,  $\hat{\sigma}_I^2$  locates the volatility clusters reasonably well. In addition, as can be seen from the absolute difference between  $\sigma^2(\xi)$  and  $\hat{\sigma}_I^2$  in Figure 8(d), the regions of higher error occur at areas of high conditional variances. Attributing this error to the realisation of the background Gaussian noise is also consistent with the fact that volatility clusters are also identified for the GMA in Figure 8(c).

To illustrate the other limitations of the method which we mentioned, we examine several outliers in Figure 7. First, we focus on the outliers in the values of  $q$  such as that corresponding to Data set 38. Figure 9 shows its MRV chart, true conditional variance surface, and estimated local variance surface for the selected  $q$  values and the median  $q$  value calculated over the 100 data sets, 23. We see that the most prominent variance cluster (i.e. the brightest spot in Plot (b) which lies away from the boundaries) do not translate into the region of highest estimated local variance (i.e. the brightest spot in Plot (d)). This means that our inference method focuses its attention on theoretically less prominent clusters, in this case those at the bottom left corner of Plot (b). Since the estimates for these clusters happen to be concentrated together without much distinction from each other, our inference method eventually groups them together to form one big cluster and selects a larger  $q$  value. This leads to a lower  $\hat{B}$  value since the overlaps in the capture window when computing  $\hat{\sigma}_I^2$  determine the amount of correlation the values have.

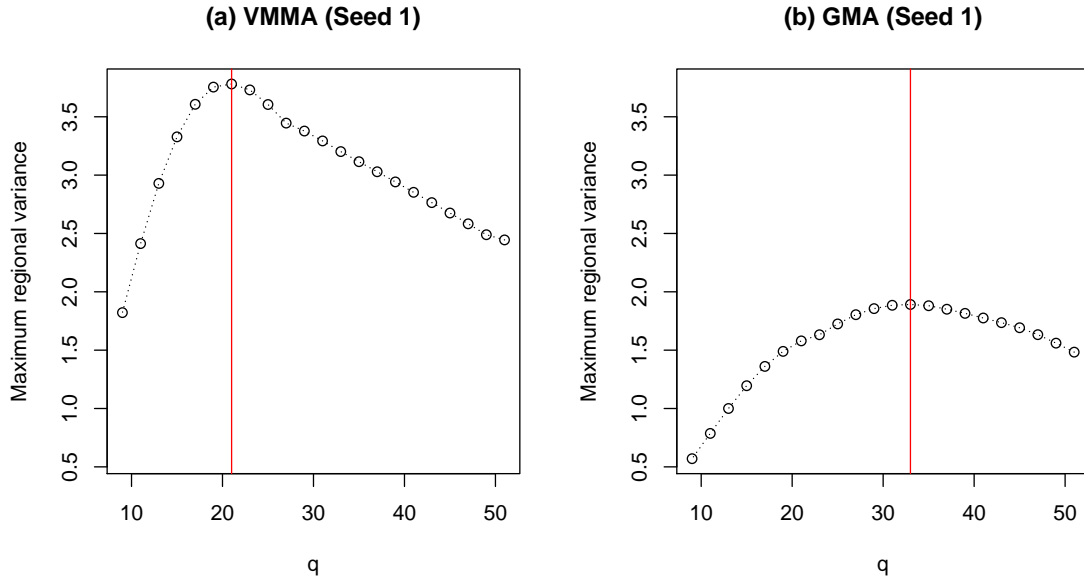


Figure 5: MRV as a function of the capture window size  $q$  for: (a) the VMMA data set; (b) the GMA data set. The red vertical lines denote the peaks of the MRVs.

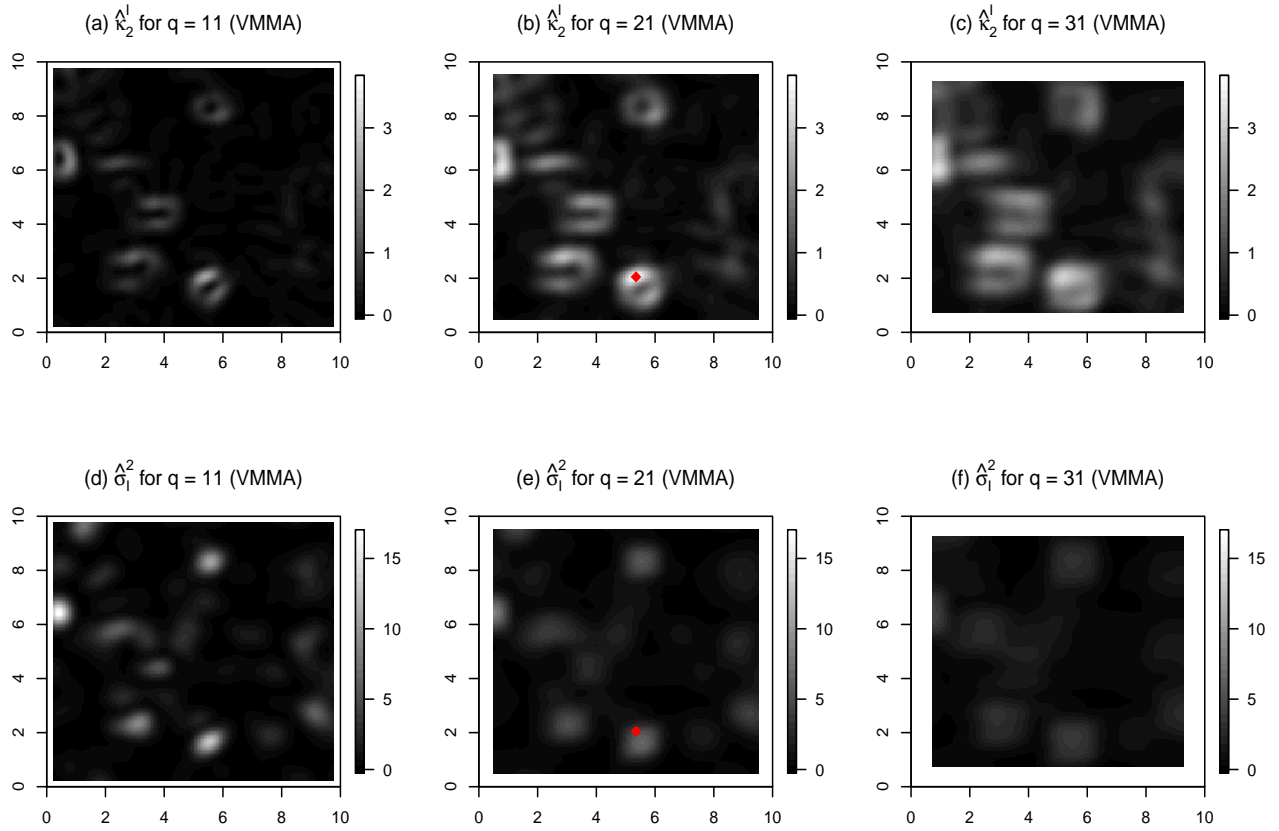


Figure 6: Data set 1: The empirical regional variance ( $\hat{\kappa}_2^I$ ) and the estimated local variance ( $\hat{\sigma}_I^2$ ) surfaces for  $q = 11, 21$  and 31. The red diamonds in Plots (b) and (e) denote the point whose regional variance is the MRV for  $q = 21$ .

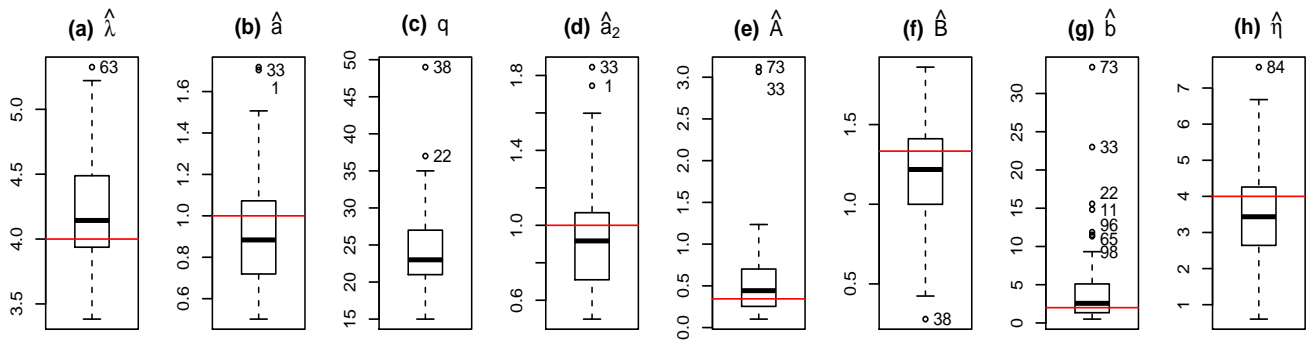


Figure 7: Box plots of estimates and selected  $q$  values from 100 VMMA data sets. The red horizontal lines denote the true parameter values and the numbers in the box plots represent the index of the data sets which give rise to the corresponding outliers.

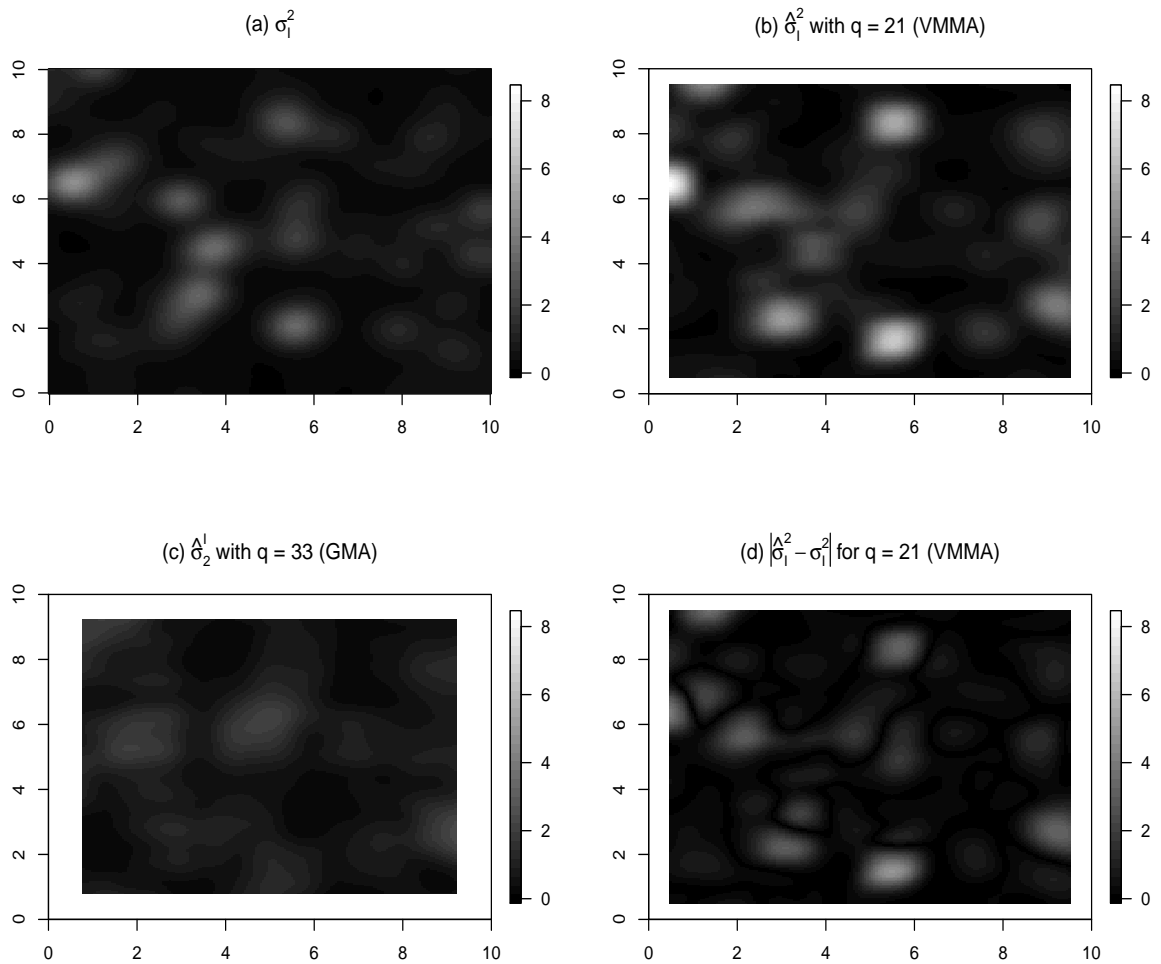


Figure 8: Data set 1: (a) The true conditional variance surface ( $\hat{\sigma}_I^2$ ); (b)-(c) the estimated local variance ( $\hat{\sigma}_I^2$ ) surfaces for the VMMA data set with  $q = 21$  and the GMA data set with  $q = 33$  respectively; and (d) the absolute difference between  $\hat{\sigma}_I^2$  and  $\hat{\sigma}_I^2$ .

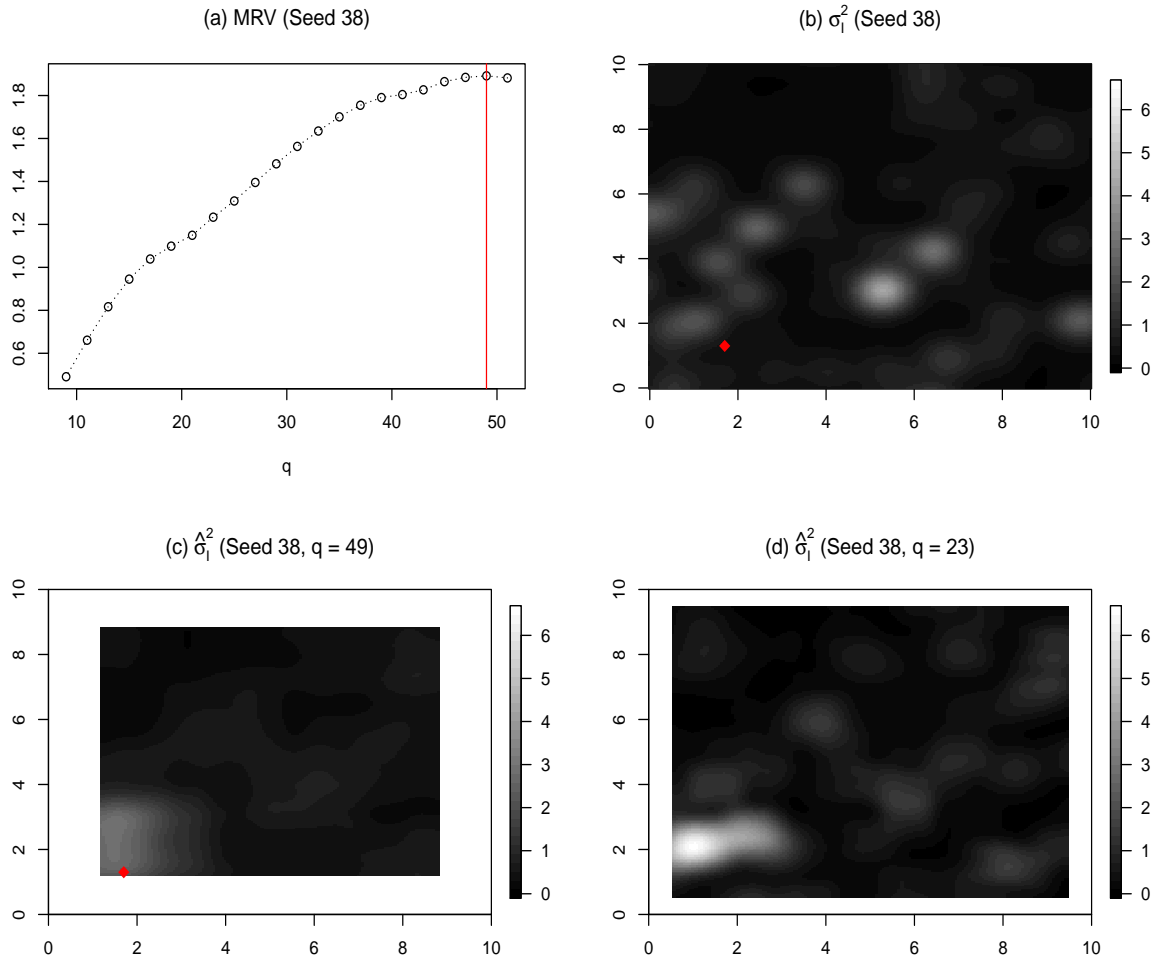


Figure 9: VMMA data set 38: (a) The MRV chart for  $q$  ranging from 9 to 51; (b) the true conditional variance ( $\hat{\sigma}_I^2$ ) surface; (c)-(d) the estimated local variance ( $\hat{\sigma}_I^2$ ) surfaces with the selected  $q$  value (49) and the median  $q$  value (23).

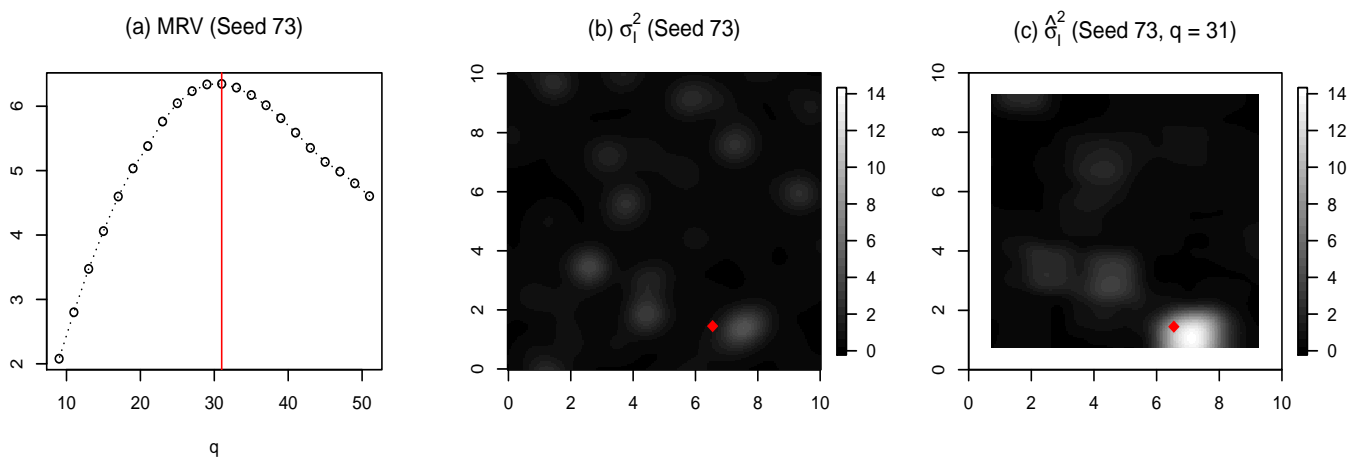


Figure 10: VMMA data set 73: (a) The MRV chart for  $q$  ranging from 9 to 51; (b) the true conditional variance ( $\hat{\sigma}_I^2$ ) surface; (c) the estimated local variance ( $\hat{\sigma}_I^2$ ) surface with the selected  $q$  value. In Plot (a), the red vertical lines denote the peaks in the MRV chart. The lighter regions in the heat plots of (b) and (c) denote higher values.



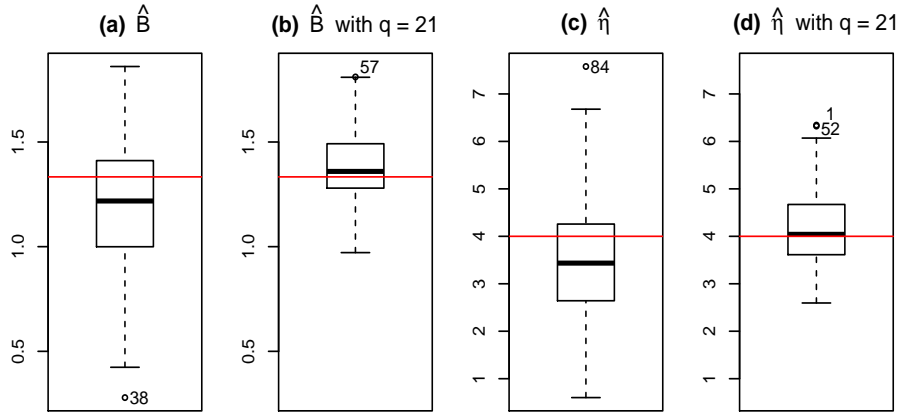


Figure 11: Box plots of estimates from 100 VMMA data sets when we set  $q = 21$ . The red horizontal lines denote the true parameter values.

The next interesting outlier corresponds to Data set 73. In this case,  $A$  which is the variance of  $\sigma_I^2$  is overestimated. Looking at the  $\hat{\sigma}_I^2$  surface in Figure 10(c), we see that this is because one sharp peak in  $\sigma_I^2$  is very distinct from the rest of the values. Due to the overestimation of  $A$ , the parameter  $b$  is also overestimated.

The two limitations mentioned are related to the multiplicative Gaussian noise in our simulations and further work is required to overcome these. Nevertheless, the potential of the estimation method, given a well-chosen  $q$  value, can be seen from the better accuracy and precision in the  $b$  and  $\eta$  estimates in Figure 11 when  $q = 21$ .

## 6 Empirical example

In the previous section, promising inference results were obtained under strong heteroskedasticity. For Model (6), this means that the parameters  $b$  and  $\eta$  are of comparable magnitude to  $\lambda$  so that there is sufficient variation in the conditional variance surface to identify the volatility clusters. In general, it is harder to estimate the parameters of  $\sigma^2$  under low heteroskedasticity. In this section, we show that even in this case, modelling the data by a VMMA instead of a GMA can be beneficial. Specifically, better prediction intervals can be obtained by using the estimated local variance surface.

We illustrate this with a data set of sea surface temperature anomalies (SSTA) for the week 29<sup>th</sup> May 2016 to 4<sup>th</sup> May 2016 (International Research Institute for Climate and Society 2016). These are calculated with respect to the 1971-2000 climatology and thus indicate how SST has changed at different spatial locations. The data, which is pictured in Figure 12(a), is given in  $^{\circ}\text{C}$  and lies on a  $1^{\circ}$  latitude/longitude grid in the Pacific Ocean between  $150.5^{\circ}\text{E}$  and  $234.5^{\circ}\text{E}$ , and  $-69.5^{\circ}\text{N}$  and  $59.5^{\circ}\text{N}$ .

Before we start our analysis, we randomly choose 100 test points away from the boundaries (denoted by the black circles in Figure 12(a)) and remove them from our data. Next, we apply median polishing on the remaining data to obtain a spatial trend. Note that this also gives trend estimates where we had missing values. The median polish algorithm has been used for various data sets in Cressie (1993). From the estimated median polish surface in Figure 12(b), we see that the trend is more prominent in the direction of the latitude and captures some of the extreme values near  $60^{\circ}\text{N}$ .

By treating  $10^{\circ}$  latitude/longitude as one unit, we fit Model (6) to the median polish residuals in Figure 12(c). All the averaging required in for example, mean and variance calculations, have been adapted to deal with the missing data. Recall that Model (6) is given by:

$$\left. \begin{aligned} Y(\mathbf{x}) &= \int_{\mathbb{R}^2} \frac{\lambda}{\pi} \exp\left(-\lambda(\mathbf{x} - \boldsymbol{\xi})^T(\mathbf{x} - \boldsymbol{\xi})\right) \sigma(\boldsymbol{\xi}) W(d\boldsymbol{\xi}), \\ \text{where } \sigma^2(\boldsymbol{\xi}) &= \int_{\mathbb{R}^2} \frac{\eta}{\pi} \exp\left(-\eta(\boldsymbol{\xi} - \mathbf{u})^T(\boldsymbol{\xi} - \mathbf{u})\right) L(d\mathbf{u}). \end{aligned} \right\}$$

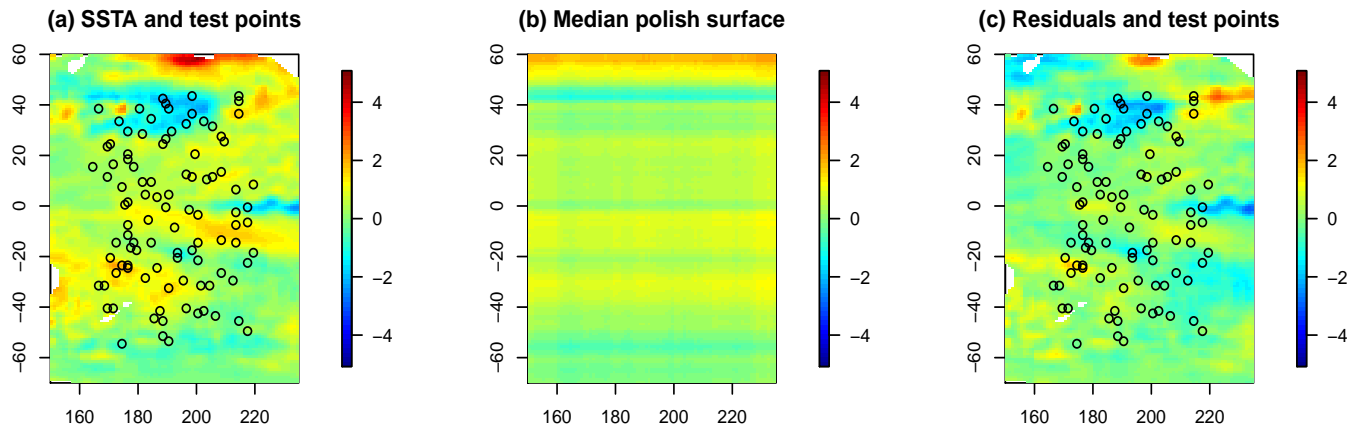


Figure 12: (a) The SSTA data set (in  $^{\circ}\text{C}$ ) where the black circles mark the 100 test points. The white regions denote missing data due to land mass; (b) the fitted median polish surface; and (c) the median polish residuals.

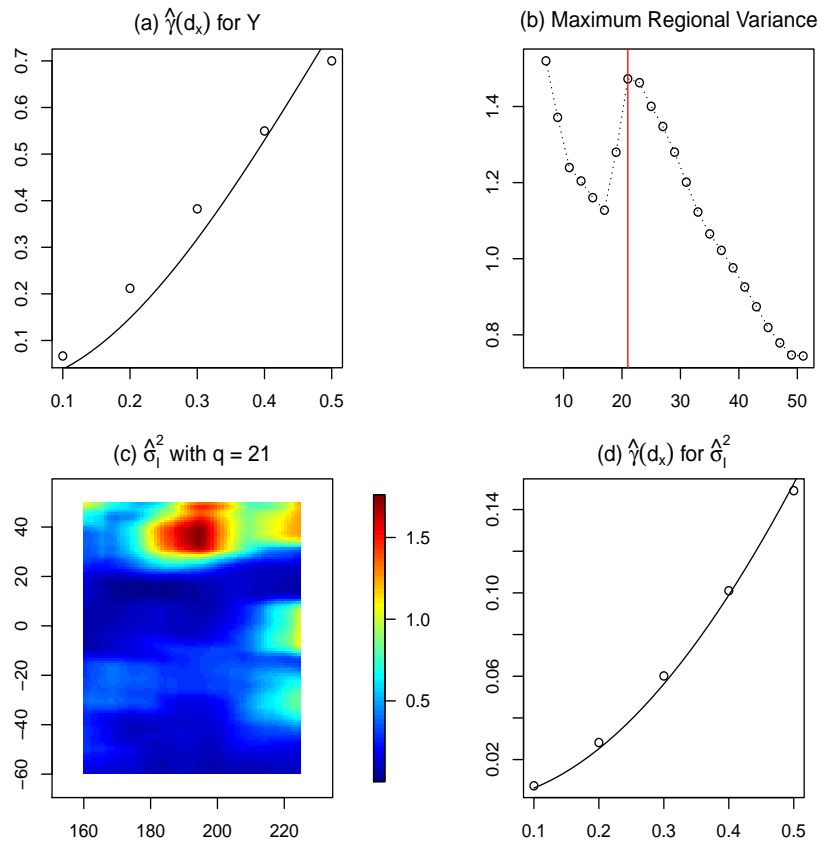


Figure 13: (a) The least-squares fit of the normalised variogram of  $Y$ ; (b) the MRV plot; (c) the estimated local variance ( $\hat{\sigma}_I^2$ ) surface with the selected  $q$  value (21); and (d) the least-squares fit of the normalised variogram of  $\hat{\sigma}_I^2$ .

Here, the rate parameters  $\lambda, \eta > 0$ ,  $W$  is a homogeneous standard Gaussian basis independent of the stochastic volatility  $\sigma(\xi)$  and  $L$  is a homogeneous subordinator basis with the mean and variance of its seed being  $a$  and  $b > 0$ . In this case, it seems reasonable to use Gaussian kernels because of tractable covariances and since they can be inferred from physical ocean dynamics (Higdon 1998, Barnes 1964).

In our simulation experiments in Section 5, we used the first spatial lag to fit the normalised variograms. For empirical data, however, better fits to the empirical normalised variograms can be obtained by fitting a curve using least-squares. Figure 13(a) shows the least squares fit for the estimated normalised variogram of the median polish residuals (i.e.  $Y$ ) when we use 5 spatial lags. The consistency of the resultant estimators can be derived by applying an edited version of Theorem 3.1 on page 70 of Lahiri et al. (2002) to our original proofs.

Based on the MRV plot in Figure 13(b), we use  $q = 21$  in the moving window step of the inference. The presence of a peak as well as the eventual decrease to the global variance in the MRV plot support the use of the VMMA for this data set. From the estimated local variance surface in Figure 13(c), we identify a prominent volatility cluster near  $40^\circ\text{N}/190^\circ\text{E}$ . This corresponds to the region in Figure 12(a) where we observe adjacent high and low anomalies.

With the estimated local variance surface, we continue our inference method. The least squares fit for the estimated normalised variogram of  $\hat{\sigma}_I^2$  is shown in Figure 13(d). This is much better than the variogram fit for  $Y$ . The final parameter estimates obtained are  $\hat{\lambda} = 3.848$ ,  $\hat{a} = 0.676$ ,  $\hat{a}_2 = 0.636$ ,  $\hat{b} = 3.499$  and  $\hat{\eta} = 0.691$ .

Despite the low value for  $\hat{\eta}$  which implies large, diffuse volatility clusters, there is still some advantage of modelling the median polish residuals by a VMMA instead of a GMA. Using the estimated local variance surface and the median polish surface, we can construct 95% confidence intervals for the SSTA values at our test points since we know that  $Y(\mathbf{x})|\mathcal{F}^\sigma \sim N(0, \sigma_I^2(\mathbf{x}))$ . At the same time, we can compute 95% confidence intervals from the corresponding GMA model for which  $\sigma_I^2(\mathbf{x}) = \hat{a}\hat{\lambda}/2\pi$  for all test points. Although the VMMA gives more narrow intervals when volatility is low (and wider intervals when volatility is high), we find that 93/100 test points lie within the confidence intervals constructed by the VMMA while only 89/100 of them lie within those constructed by the GMA. The benefit of modelling with a VMMA is expected to be enhanced under stronger heteroskedastic behaviour.

## 7 Conclusion and further work

In this paper, we focused on volatility modulated moving averages (VMMA) and their ability to exhibit spatial heteroskedasticity. These processes extend the definition of a Gaussian moving average or process convolution by introducing a stationary stochastic volatility field in the integral.

After providing a summary of how such an integral is constructed in Section 2, we derived several distributional properties of a VMMA in Section 3. These were used to develop a two-step moments-matching estimation procedure in Section 5. In the first step, we relied on the stationarity of a VMMA together with its second-order properties; in the second step, we examined the conditional non-stationarity in the form of local variances and used the second-order properties of the conditional variance process. Consistency of the resulting estimators can be proved under suitable double asymptotics.

In Section 4, we developed a discrete convolution simulation algorithm for VMMA and provided semi-explicit formulas for the mean squared error. We also derived an explicit formula for an upper bound which proved to be more useful in practice for deriving orders of convergence. Through experiments with simulated data, we tested our inference procedure and found that promising results were achieved under strong heteroskedasticity. As expected, the outlier analysis also revealed a small degree of sensitivity to the realisation of the multiplicative Gaussian noise in our VMMA.

The application to sea surface temperature anomaly data in Section 6 illustrates the benefits of using a VMMA instead of a GMA when spatial heteroskedasticity is present. Better prediction for missing values can be achieved through the estimated trend and the estimated local variances. This empirical example also shines light on the many ways one can adapt and improve the two-step moments-matching estimation method. Using least-squares in the fitting of the normalised variograms is one such extension. To further improve the variogram fits, generalising the method to the popular Matérn kernel as well as anisotropic kernels will be helpful. One may also consider experimenting with different shapes of the moving window to capture that of the anisotropic volatility clusters. Incorporating kernel averaging in the moving win-

dow approach, such as that done for the local variogram estimation in Fouedjio et al. (2016), could also help to reduce the occurrence of overlapping clusters.

So far, only point estimates of our parameters are considered and the uncertainty in these is not reflected in our prediction intervals. A key direction for further work is building a Bayesian framework around our moments-matching method to provide credible intervals. In this case, the moments-matching estimates could be useful as starting values or to motivate prior distributions. Composite likelihoods could also be used to reduce computational effort.

## Appendix:

*Proof of Example 7.* For Model (6), we have  $\text{Cov}(\sigma^2(\xi), \sigma^2(\xi^*)) = \frac{b\eta}{2\pi} \exp\left(-\frac{\eta(\xi - \xi^*)^T(\xi - \xi^*)}{2}\right)$ , where  $\text{Var}(L') = b$ . So, with  $\mathbf{w} = \mathbf{x} - \xi$ ,  $\mathbf{w}^* = \mathbf{x}^* - \xi^*$  and  $\mathbf{u} = \mathbf{x} - \mathbf{x}^*$ :

$$\begin{aligned} \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x}^*)) &= \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) g^2(\mathbf{x}^* - \xi^*) \text{Cov}(\sigma^2(\xi), \sigma^2(\xi^*)) d\xi d\xi^* \\ &= \frac{b\lambda^4\eta}{2\pi^5} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \exp\left(-2\lambda[\mathbf{w}^T\mathbf{w} + \mathbf{w}^{*T}\mathbf{w}^*] - \frac{\eta(\mathbf{w}^* - \mathbf{w} + \mathbf{u})^T(\mathbf{w}^* - \mathbf{w} + \mathbf{u})}{2}\right) d\mathbf{w} d\mathbf{w}^*. \end{aligned}$$

The exponent of the integrand is equal to:

$$-2\lambda(w_1^2 + w_1^{*2}) - \frac{\eta}{2}[(w_1^* - w_1)^2 + 2u_1(w_1^* - w_1)] - 2\lambda(w_2^2 + w_2^{*2}) - \frac{\eta}{2}[(w_2^* - w_2)^2 + 2u_2(w_2^* - w_2)] - \frac{\eta}{2}(u_1^2 + u_2^2).$$

Focusing on the terms in  $w_1$  and  $w_1^*$ :

$$\begin{aligned} &\int_{\mathbb{R}} \int_{\mathbb{R}} \exp\left(-2\lambda(w_1^2 + w_1^{*2}) - \frac{\eta}{2}[(w_1^* - w_1)^2 + 2u_1(w_1^* - w_1)]\right) dw_1 dw_1^* \\ &= \int_{\mathbb{R}} \exp\left(-\left[2\lambda + \frac{\eta}{2}\right]w_1^{*2} - \eta u_1 w_1^* + \frac{\eta[u_1 + w_1^*]^2}{4(2\lambda + \frac{\eta}{2})}\right) \int_{\mathbb{R}} \exp\left(-\left[2\lambda + \frac{\eta}{2}\right]\left[w_1 - \frac{\eta[u_1 + w_1^*]}{2(2\lambda + \frac{\eta}{2})}\right]^2\right) dw_1 dw_1^* \\ &= \sqrt{\frac{\pi}{2\lambda + \frac{\eta}{2}}} \int_{\mathbb{R}} \exp\left(-\left[2\lambda + \frac{\eta}{2}\right]w_1^{*2} - \eta u_1 w_1^* + \frac{\eta[u_1 + w_1^*]^2}{4(2\lambda + \frac{\eta}{2})}\right) dw_1^* \\ &= \sqrt{\frac{\pi}{2\lambda + \frac{\eta}{2}}} \exp\left(\frac{\eta^2 u_1^2}{4(2\lambda + \frac{\eta}{2})} + \frac{\lambda(4\lambda + 2\eta)}{2\lambda + \frac{\eta}{2}} \left[\frac{\eta u_1}{4\lambda + 2\eta}\right]^2\right) \int_{\mathbb{R}} \exp\left(-\frac{\lambda(4\lambda + 2\eta)}{2\lambda + \frac{\eta}{2}} \left[w_1^* - \frac{\eta u_1}{4\lambda + 2\eta}\right]^2\right) dw_1^* \\ &= \frac{\pi}{\sqrt{\lambda(4\lambda + 2\eta)}} \exp\left(\frac{\eta^2 u_1^2}{4\lambda + 2\eta}\right). \end{aligned}$$

As the terms in  $w_2$  and  $w_2^*$  follow a similar structure, we have:

$$\text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x}^*)) = \frac{b\lambda^3\eta}{4\pi^3(2\lambda + \eta)} \exp\left(\frac{-\lambda\eta}{2\lambda + \eta}(\mathbf{x} - \mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*)\right). \quad (17)$$

On the other hand:

$$\begin{aligned} \mathbb{E}\left[\left(\int_{\mathbb{R}^2} g(\mathbf{x} - \xi) g(\mathbf{x}^* - \xi) \sigma^2(\xi) d\xi\right)^2\right] &= \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) g(\mathbf{x}^* - \xi) g(\mathbf{x} - \xi^*) g(\mathbf{x}^* - \xi^*) \mathbb{E}[\sigma^2(\xi) \sigma^2(\xi^*)] d\xi d\xi^* \\ &= \frac{\lambda^4}{\pi^4} \left(\frac{b\eta A'}{2\pi} + a^2 B'\right), \end{aligned}$$

where  $A' = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} e^{-\lambda[\mathbf{w}^T\mathbf{w} + \mathbf{w}^{*T}\mathbf{w}^* + (\mathbf{w}^* + \mathbf{u})^T(\mathbf{w}^* + \mathbf{u}) + (\mathbf{w} - \mathbf{u})^T(\mathbf{w} - \mathbf{u})] - \frac{\eta}{2}(\mathbf{w}^* - \mathbf{w} + \mathbf{u})^T(\mathbf{w}^* - \mathbf{w} + \mathbf{u})} d\mathbf{w} d\mathbf{w}^*$ , and  $B' = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} e^{-\lambda[\mathbf{w}^T\mathbf{w} + \mathbf{w}^{*T}\mathbf{w}^* + (\mathbf{w}^* + \mathbf{u})^T(\mathbf{w}^* + \mathbf{u}) + (\mathbf{w} - \mathbf{u})^T(\mathbf{w} - \mathbf{u})]} d\mathbf{w} d\mathbf{w}^*$ .

The exponent of the  $A'$ 's integrand is equal to:

$$\begin{aligned} & -2\lambda [w_1^2 + w_1^{*2} + u_1 w_1^* - u_1 w_1] - \frac{\eta}{2} [(w_1^* - w_1)^2 + 2u_1 (w_1^* - w_1)] - \left[2\lambda + \frac{\eta}{2}\right] u_1^2 \\ & -2\lambda [w_2^2 + w_2^{*2} + u_2 w_2^* - u_2 w_2] - \frac{\eta}{2} [(w_2^* - w_2)^2 + 2u_2 (w_2^* - w_2)] - \left[2\lambda + \frac{\eta}{2}\right] u_2^2. \end{aligned}$$

Focusing on the terms in  $w_1$  and  $w_1^*$ :

$$\begin{aligned} & \int_{\mathbb{R}} \int_{\mathbb{R}} \exp \left( -2\lambda [w_1^2 + w_1^{*2} + u_1 w_1^* - u_1 w_1] - \frac{\eta}{2} [(w_1^* - w_1)^2 + 2u_1 (w_1^* - w_1)] \right) dw_1 dw_1^* \\ & = \int_{\mathbb{R}} \exp \left( -\left[2\lambda + \frac{\eta}{2}\right] w_1^{*2} - [2\lambda + \eta] u_1 w_1^* \right) \int_{\mathbb{R}} \exp \left( -\left[2\lambda + \frac{\eta}{2}\right] w_1^2 + [(2\lambda + \eta) u_1 + \eta w_1^*] w_1 \right) dw_1 dw_1^* \\ & = \sqrt{\frac{\pi}{2\lambda + \frac{\eta}{2}}} \exp \left( \frac{(2\lambda + \eta) u_1^2}{2} \right) \int_{\mathbb{R}} \exp \left( -\frac{4\lambda^2 + 2\lambda\eta}{2\lambda + \frac{\eta}{2}} \left[ w_1 - \frac{u_1}{2} \right]^2 \right) dw_1^* \\ & = \frac{\pi}{\sqrt{4\lambda^2 + 2\lambda\eta}} \exp \left( \frac{(2\lambda + \eta) u_1^2}{2} \right). \end{aligned}$$

Since the terms in  $w_2$  and  $w_2^*$  follow the same form:

$$A' = \frac{\pi^2}{4\lambda^2 + 2\lambda\eta} \exp \left( \left[ \frac{(2\lambda + \eta)}{2} - \left( 2\lambda + \frac{\eta}{2} \right) \right] (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \right) = \frac{\pi^2}{4\lambda^2 + 2\lambda\eta} \exp \left( -\lambda (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \right).$$

The exponent of the  $B'$ 's integrand is equal to:

$$-2\lambda [(w_1^2 - u_1 w_1) + (w_1^{*2} + u_1 w_1^*) + (w_2^2 - u_2 w_2) + (w_2^{*2} + u_2 w_2^*) + [u_1^2 + u_2^2]].$$

Focusing on the terms in  $w_1$ :

$$\int_{\mathbb{R}} \exp (-2\lambda (w_1^2 - u_1 w_1)) dw_1 = \exp \left( \frac{\lambda u_1^2}{2} \right) \int_{\mathbb{R}} \exp \left( -2\lambda \left( w_1 - \frac{u_1}{2} \right)^2 \right) dw_1 = \sqrt{\frac{\pi}{2\lambda}} \exp \left( \frac{\lambda u_1^2}{2} \right).$$

Similarly,  $\int_{\mathbb{R}} \exp (-2\lambda (w_1^{*2} + u_1 w_1^*)) = \sqrt{\frac{\pi}{2\lambda}} \exp \left( \frac{\lambda u_1^2}{2} \right)$ . Thus,  $B' = \frac{\pi^2}{4\lambda^2} \exp \left( -\lambda (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \right)$ .

Since  $\text{Cov}(Y^2(\mathbf{x}), Y^2(\mathbf{x}^*)) = \frac{2\lambda^4}{\pi^4} \left( \frac{b\eta A'}{2\pi} + a^2 B' \right) + \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x}^*))$ , we obtain the required result.  $\square$

*Proof of Theorem 3.*

$$\begin{aligned} \mathbb{E} [|Y(\mathbf{x}) - Z(\mathbf{x})|^2] &= \mathbb{E} \left[ \left| \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) \sigma(\xi) W(d\xi) - \int_{\mathbb{R}^2} g_{\Delta}(\mathbf{x}, \xi) \sigma_{\Delta}(\xi) W(d\xi) \right|^2 \right] \\ &= \mathbb{E} \left[ \left| \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) (\sigma(\xi) - \sigma_{\Delta}(\xi)) W(d\xi) + \int_{\mathbb{R}^2} (g(\mathbf{x} - \xi) - g_{\Delta}(\mathbf{x}, \xi)) \sigma_{\Delta}(\xi) W(d\xi) \right|^2 \right] \\ &= T1 + T2 + T3, \end{aligned}$$

where  $T1 := \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) \mathbb{E} [(\sigma(\xi) - \sigma_{\Delta}(\xi))^2] d\xi$ ,  $T2 := \int_{\mathbb{R}^2} (g(\mathbf{x} - \xi) - g_{\Delta}(\mathbf{x}, \xi))^2 \mathbb{E} [\sigma_{\Delta}^2(\xi)] d\xi$  and  $T3 := 2 \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) (g(\mathbf{x} - \xi) - g_{\Delta}(\mathbf{x}, \xi)) \mathbb{E} [\sigma_{\Delta}(\xi) (\sigma(\xi) - \sigma_{\Delta}(\xi))] d\xi$ . In the calculations, the third equality follows from the independence of  $\sigma^2$  and  $W$  as well as the fact that  $W$  is a homogeneous standard Gaussian basis. We simplify the three terms separately:

$$\begin{aligned} T1 &= \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) \mathbb{E} [\sigma^2(\xi)] d\xi + \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) \mathbb{E} [\sigma_{\Delta}^2(\xi)] d\xi - 2 \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) \mathbb{E} [\sigma(\xi) \sigma_{\Delta}(\xi)] d\xi \\ &= a \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi \right) \left( \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} + \sum_{i,j=-\tilde{p}}^{\tilde{p}} h(i\Delta, j\Delta) \Delta^2 \right) - 2 \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) \mathbb{E} [\sigma(\xi) \sigma_{\Delta}(\xi)] d\xi \end{aligned}$$

since  $\mathbb{E} [\sigma^2(\xi)] = \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) \mathbb{E} [L(d\mathbf{u})] = a \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u}$ , and:

$$\mathbb{E} [\sigma_{\Delta}^2(\xi)] = \int_{\mathbb{R}^2} h_{\Delta}(\xi, \mathbf{u}) \mathbb{E} [L(d\mathbf{u})] = a \sum_{i', j' = -\tilde{p}}^{\tilde{p}} h(i' \Delta, j' \Delta) \Delta^2.$$

Next, we focus on  $T2$ :

$$\begin{aligned} T2 &= \left( a \sum_{i', j' = -\tilde{p}}^{\tilde{p}} h(i' \Delta, j' \Delta) \Delta^2 \right) \left[ \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi + \int_{\mathbb{R}^2} g_{\Delta}^2(\mathbf{x}, \xi) d\xi - 2 \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) g_{\Delta}(\mathbf{x}, \xi) d\xi \right] \\ &= \left[ \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi - \sum_{i, j = -p}^p g^2(i \Delta, j \Delta) \Delta^2 \right) \right. \\ &\quad \left. + 2 \sum_{i, j = -p}^p g(i \Delta, j \Delta) \left( g(i \Delta, j \Delta) \Delta^2 - \int_{i \Delta - \frac{\Delta}{2}}^{i \Delta + \frac{\Delta}{2}} \int_{j \Delta - \frac{\Delta}{2}}^{j \Delta + \frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right] \times \left( a \sum_{i', j' = -\tilde{p}}^{\tilde{p}} h(i' \Delta, j' \Delta) \Delta^2 \right), \end{aligned}$$

by letting  $\mathbf{w} = \xi - \mathbf{x}$  and since:

$$\begin{aligned} \int_{\mathbb{R}^2} g_{\Delta}^2(\mathbf{x}, \xi) d\xi &= \sum_{i, j = -p}^p \sum_{i', j' = -p}^p \int_{\mathbb{R}^2} \mathbf{1}_{[x_1 + i \Delta - \frac{\Delta}{2}, x_1 + i \Delta + \frac{\Delta}{2})} (s_1) \mathbf{1}_{[x_2 + j \Delta - \frac{\Delta}{2}, x_2 + j \Delta + \frac{\Delta}{2})} (s_2) \\ &\quad \mathbf{1}_{[x_1 + i' \Delta - \frac{\Delta}{2}, x_1 + i' \Delta + \frac{\Delta}{2})} (s_1) \mathbf{1}_{[x_2 + j' \Delta - \frac{\Delta}{2}, x_2 + j' \Delta + \frac{\Delta}{2})} (s_2) g(i \Delta, j \Delta) g(i' \Delta, j' \Delta) d\xi \\ &= \sum_{i, j = -p}^p g^2(i \Delta, j \Delta) \int_{\mathbb{R}^2} \mathbf{1}_{[x_1 + i \Delta - \frac{\Delta}{2}, x_1 + i \Delta + \frac{\Delta}{2})} (s_1) \mathbf{1}_{[x_2 + j \Delta - \frac{\Delta}{2}, x_2 + j \Delta + \frac{\Delta}{2})} (s_2) d\xi \\ &= \sum_{i, j = -p}^p g^2(i \Delta, j \Delta) \Delta^2, \end{aligned}$$

and

$$\begin{aligned} \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) g_{\Delta}(\mathbf{x}, \xi) d\xi &= \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) \sum_{i, j = -p}^p \mathbf{1}_{[x_1 + i \Delta - \frac{\Delta}{2}, x_1 + i \Delta + \frac{\Delta}{2})} (s_1) \mathbf{1}_{[x_2 + j \Delta - \frac{\Delta}{2}, x_2 + j \Delta + \frac{\Delta}{2})} (s_2) \\ &\quad \times g(i \Delta, j \Delta) d\xi \\ &= \sum_{i, j = -p}^p g(i \Delta, j \Delta) \int_{i \Delta - \frac{\Delta}{2}}^{i \Delta + \frac{\Delta}{2}} \int_{j \Delta - \frac{\Delta}{2}}^{j \Delta + \frac{\Delta}{2}} g(\mathbf{w}) dw_1 dw_2. \end{aligned}$$

Finally, we look at  $T3$ :

$$\begin{aligned} T3 &= 2 \int_{\mathbb{R}^2} [g^2(\mathbf{x} - \xi) - g(\mathbf{x} - \xi) g_{\Delta}(\mathbf{x}, \xi)] (\mathbb{E} [\sigma(\xi) \sigma_{\Delta}(\xi)] - \mathbb{E} [\sigma_{\Delta}^2(\xi)]) d\xi \\ &= 2 \left[ \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) \mathbb{E} [\sigma(\xi) \sigma_{\Delta}(\xi)] d\xi - \int_{\mathbb{R}^2} g(\mathbf{x} - \xi) g_{\Delta}(\mathbf{x}, \xi) \mathbb{E} [\sigma(\xi) \sigma_{\Delta}(\xi)] d\xi \right. \\ &\quad \left. - a \sum_{i', j' = -\tilde{p}}^{\tilde{p}} h(i' \Delta, j' \Delta) \Delta^2 \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi - \sum_{i, j = -p}^p g(i \Delta, j \Delta) \int_{i \Delta - \frac{\Delta}{2}}^{i \Delta + \frac{\Delta}{2}} \int_{j \Delta - \frac{\Delta}{2}}^{j \Delta + \frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right]. \end{aligned}$$

To obtain the required expression for  $\mathbb{E} [\sigma(\xi) \sigma_{\Delta}(\xi)]$ , we apply the following equality from Section 1.7 of Applebaum (2009):  $u^{\alpha} = \frac{\alpha}{\Gamma(1-\alpha)} \int_0^{\infty} (1 - e^{-ux}) \frac{dx}{x^{1+\alpha}}$ , where  $u \geq 0$  and  $0 < \alpha < 1$ . By setting  $u = \sigma(\xi)$  and  $u = \sigma_{\Delta}(\xi)$  separately with  $\alpha = 1/2$ , and using Fubini's Theorem:

$$\begin{aligned} \mathbb{E} [\sigma(\xi) \sigma_{\Delta}(\xi)] &= \mathbb{E} \left[ \frac{1}{4\pi} \int_0^{\infty} \int_0^{\infty} (1 - e^{-\sigma^2(\xi)x}) (1 - e^{-\sigma_{\Delta}^2(\xi)y}) \frac{dx}{x^{3/2}} \frac{dy}{y^{3/2}} \right] \\ &= \frac{1}{4\pi} \int_0^{\infty} \int_0^{\infty} \left( 1 - \mathbb{E} [e^{-\sigma^2(\xi)x}] - \mathbb{E} [e^{-\sigma_{\Delta}^2(\xi)y}] + \mathbb{E} [e^{-\sigma^2(\xi)x - \sigma_{\Delta}^2(\xi)y}] \right) \frac{dx}{x^{3/2}} \frac{dy}{y^{3/2}}. \end{aligned}$$

Since  $\sigma^2(\xi) = \int_{\mathbb{R}^2} h(\xi - \mathbf{u})L(d\mathbf{u})$ , its CGF can be expressed as:  $C(\theta; \sigma^2(\xi)) = \int_{\mathbb{R}^2} C(\theta h(\xi - \mathbf{u}); L')d\mathbf{u}$ .

By replacing  $\theta$  by  $ix$ , we find that the Laplace exponent of  $\sigma^2(\xi)$  is equal to  $\int_{\mathbb{R}^2} \Psi_L(xh(\xi - \mathbf{u}))d\mathbf{u}$ . Thus,  $\mathbb{E} \left[ e^{-\sigma^2(\xi)x} \right] = e^{\int_{\mathbb{R}^2} \Psi_L(xh(\xi - \mathbf{u}))d\mathbf{u}}$ . The expressions for  $\mathbb{E} \left[ e^{-\sigma_\Delta^2(\xi)y} \right]$  and  $\mathbb{E} \left[ e^{-\sigma^2(\xi)x - \sigma_\Delta^2(\xi)y} \right]$  can be found analogously.  $\square$

*Proof of Corollary 3.* We obtain an upper bound for  $T1$  in Theorem 3. By using the fact that the harmonic mean of  $\sigma^2$  and  $\sigma_\Delta^2$  is less than or equal to their geometric mean, and by applying Jensen's inequality since  $1/x$  is a convex function of  $x$  for  $x > 0$ :

$$\begin{aligned} \mathbb{E} [\sigma(\xi)\sigma_\Delta(\xi)] &\geq \mathbb{E} \left[ 2 \left( \frac{1}{\sigma^2(\xi)} + \frac{1}{\sigma_\Delta^2(\xi)} \right)^{-1} \right] \geq 2 \left( \mathbb{E} \left[ \frac{1}{\sigma^2(\xi)} \right] + \mathbb{E} \left[ \frac{1}{\sigma_\Delta^2(\xi)} \right] \right)^{-1} \geq 2 \left( \frac{1}{\mathbb{E} [\sigma^2(\xi)]} + \frac{1}{\mathbb{E} [\sigma_\Delta^2(\xi)]} \right)^{-1} \\ &\Rightarrow T1 \leq \left( a \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi \right) \left[ \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} + \sum_{i,j=-\bar{p}}^{\bar{p}} h(i\Delta, j\Delta) \Delta^2 \right. \\ &\quad \left. - 4 \left( \frac{1}{\int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u}} + \frac{1}{\sum_{i=-\bar{p}}^{\bar{p}} \sum_{j=-\bar{p}}^{\bar{p}} h(i\Delta, j\Delta) \Delta^2} \right)^{-1} d\xi \right] \\ &= \left( a \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi \right) \frac{\left( \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} - \sum_{i=-\bar{p}}^{\bar{p}} \sum_{j=-\bar{p}}^{\bar{p}} h(i\Delta, j\Delta) \Delta^2 \right)^2}{\int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} + \sum_{i=-\bar{p}}^{\bar{p}} \sum_{j=-\bar{p}}^{\bar{p}} h(i\Delta, j\Delta) \Delta^2} \\ &:= T4. \end{aligned}$$

Likewise, we obtain an upper bound for  $T3$  in Theorem 3. By using the lower bound for  $\mathbb{E} [\sigma(\xi)\sigma_\Delta(\xi)]$  attained previously as well as the fact that the arithmetic mean of  $\sigma^2$  and  $\sigma_\Delta^2$  is greater than or equal to their geometric mean, we obtain:

$$\begin{aligned} T3 &\leq 2 \left[ \frac{a}{2} \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) \left[ \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} + \sum_{i,j=-\bar{p}}^{\bar{p}} h(i\Delta, j\Delta) \Delta^2 \right] d\xi \right. \\ &\quad \left. - 2a \sum_{i,j=-p}^p g(i\Delta, j\Delta) \int_{x_1+i\Delta-\frac{\Delta}{2}}^{x_1+i\Delta+\frac{\Delta}{2}} \int_{x_2+j\Delta-\frac{\Delta}{2}}^{x_2+j\Delta+\frac{\Delta}{2}} g(\mathbf{x} - \xi) \left( \frac{1}{\int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u}} + \frac{1}{\sum_{i',j'=-\bar{p}}^{\bar{p}} h(i'\Delta, j'\Delta) \Delta^2} \right)^{-1} d\xi \right. \\ &\quad \left. - a \sum_{i',j'=-\bar{p}}^{\bar{p}} h(i'\Delta, j'\Delta) \Delta^2 \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi - \sum_{i,j=-p}^p g(i\Delta, j\Delta) \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right] \\ &= a \left[ \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} - \sum_{i,j=-\bar{p}}^{\bar{p}} h(i\Delta, j\Delta) \Delta^2 \right] \left[ \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi - \frac{2 \sum_{i',j'=-\bar{p}}^{\bar{p}} h(i'\Delta, j'\Delta) \Delta^2}{\sum_{i',j'=-\bar{p}}^{\bar{p}} h(i'\Delta, j'\Delta) \Delta^2 + \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u}} \right. \\ &\quad \left. \left( \sum_{i,j=-p}^p g(i\Delta, j\Delta) \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right] \\ &= a \left[ \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} - \sum_{i,j=-\bar{p}}^{\bar{p}} h(i\Delta, j\Delta) \Delta^2 \right] \left[ \left( \int_{\mathbb{R}^2} g^2(\mathbf{x} - \xi) d\xi - \sum_{i,j=-p}^p g^2(i\Delta, j\Delta) \Delta^2 \right) \right. \\ &\quad \left. + \sum_{i,j=-p}^p g(i\Delta, j\Delta) \left( g(i\Delta, j\Delta) \Delta^2 - \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right. \\ &\quad \left. + \left( \int_{\mathbb{R}^2} h(\xi - \mathbf{u}) d\mathbf{u} - \sum_{i',j'=-\bar{p}}^{\bar{p}} h(i'\Delta, j'\Delta) \Delta^2 \right) \left( \sum_{i,j=-p}^p g(i\Delta, j\Delta) \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w} \right) \right] \\ &:= T5. \end{aligned}$$

By combining the upper bounds of  $T1$  and  $T3$  with  $T2$ , we obtain an upper bound for the simulation MSE.  $\square$

*Proof of Lemma 1.* Using Assumption 1, we can apply second-order Taylor expansions of  $g^2(\mathbf{w})$  around  $(i\Delta, j\Delta)$  for  $-\lfloor R/\Delta \rfloor \leq i, j \leq \lfloor R/\Delta \rfloor$ :

$$\begin{aligned} & \int_{-(R+\Delta/2)}^{R+\Delta/2} \int_{-(R+\Delta/2)}^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w} - \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} g^2(i\Delta, j\Delta) \Delta^2 \\ &= \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} \int_{i\Delta-\Delta/2}^{i\Delta+\Delta/2} \int_{j\Delta-\Delta/2}^{j\Delta+\Delta/2} (g^2(\mathbf{w}) - g^2(i\Delta, j\Delta)) d\mathbf{w} \\ &= \frac{1}{2} \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} \left( \int_{i\Delta-\Delta/2}^{i\Delta+\Delta/2} \int_{j\Delta-\Delta/2}^{j\Delta+\Delta/2} \left[ (w_1 - i\Delta)^2 \frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_1} \Big|_{\mathbf{w}=\zeta_1(i,j,\Delta)} + (w_2 - j\Delta)^2 \frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_2} \Big|_{\mathbf{w}=\zeta_2(i,j,\Delta)} \right] d\mathbf{w} \right), \end{aligned}$$

where  $\zeta_k(i, j, \Delta) \in I(i, j, \Delta) = (i\Delta - \Delta/2, i\Delta + \Delta/2) \times (j\Delta - \Delta/2, j\Delta + \Delta/2)$  for  $k = 1, 2$ , and since the terms involving single powers of  $(w_1 - i\Delta)$  and  $(w_2 - j\Delta)$  integrate to zero.

Let  $s_1(i, j, \Delta)$  and  $S_1(i, j, \Delta)$  be the infimum and supremum of  $\frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_1}$  over  $I(i, j, \Delta)$ . Then:

$$\begin{aligned} \frac{\Delta^4}{12} \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} s_1(i, j, \Delta) &\leq \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} \int_{i\Delta-\Delta/2}^{i\Delta+\Delta/2} \int_{j\Delta-\Delta/2}^{j\Delta+\Delta/2} (w_1 - i\Delta)^2 \frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_1} \Big|_{\mathbf{w}=\zeta_1(i,j,\Delta)} d\mathbf{w} \\ &\leq \frac{\Delta^4}{12} \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} S_1(i, j, \Delta), \end{aligned}$$

since  $\int_{i\Delta-\Delta/2}^{i\Delta+\Delta/2} \int_{j\Delta-\Delta/2}^{j\Delta+\Delta/2} (w_1 - i\Delta)^2 d\mathbf{w} = \Delta \left[ \frac{\Delta^3}{12} \right] = \frac{\Delta^4}{12}$ .

Since  $\lim_{\Delta \rightarrow 0} \Delta^2 \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} s_1(i, j, \Delta) = \lim_{\Delta \rightarrow 0} \Delta^2 \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} S_1(i, j, \Delta) = \int_{-R}^R \int_{-R}^R \frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_1} d\mathbf{w}$ :

$$\lim_{\Delta \rightarrow 0} \frac{1}{\Delta^2} \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} \int_{i\Delta-\Delta/2}^{i\Delta+\Delta/2} \int_{j\Delta-\Delta/2}^{j\Delta+\Delta/2} (w_1 - i\Delta)^2 \frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_1} \Big|_{\mathbf{w}=\zeta_1(i,j,\Delta)} d\mathbf{w} = \frac{1}{12} \int_{-R}^R \int_{-R}^R \frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_1} d\mathbf{w},$$

by the Sandwich Theorem. Using similar arguments for  $\sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} \int_{i\Delta-\Delta/2}^{i\Delta+\Delta/2} \int_{j\Delta-\Delta/2}^{j\Delta+\Delta/2} (w_2 - j\Delta)^2 \frac{\partial^2 g^2(\mathbf{w})}{\partial^2 w_2} \Big|_{\mathbf{w}=\zeta_2(i,j,\Delta)} d\mathbf{w}$ :

$$\int_{-(R+\Delta/2)}^{R+\Delta/2} \int_{-(R+\Delta/2)}^{R+\Delta/2} g^2(\mathbf{w}) d\mathbf{w} - \sum_{i,j=-\lfloor R/\Delta \rfloor}^{\lfloor R/\Delta \rfloor} g^2(i\Delta, j\Delta) \Delta^2 = O(\Delta^2).$$

By analogous arguments, one obtain the results for  $g(i\Delta, j\Delta) \Delta^2 - \int_{i\Delta-\frac{\Delta}{2}}^{i\Delta+\frac{\Delta}{2}} \int_{j\Delta-\frac{\Delta}{2}}^{j\Delta+\frac{\Delta}{2}} g(\mathbf{w}) d\mathbf{w}$  and

$$\int_{-(\tilde{R}+\Delta/2)}^{\tilde{R}+\Delta/2} \int_{-(\tilde{R}+\Delta/2)}^{\tilde{R}+\Delta/2} h(\mathbf{w}) d\mathbf{w} - \sum_{i,j=-\tilde{R}/\Delta}^{\tilde{R}/\Delta} h(i\Delta, j\Delta) \Delta^2. \quad \square$$

*Proof of Theorem 4.* With  $\mathbb{E}[\cdot | \mathcal{F}^\sigma]$  denoting the expectation conditional on  $\sigma^2$ , the MSE of our estimator is:

$$\begin{aligned} & \mathbb{E} \left[ \left| \frac{1}{(2Q+1)^2} \sum_{l=-Q}^Q \sum_{k=-Q}^Q Y^2(\mathbf{x} + (l, k)\Delta) - \sigma_I^2(\mathbf{x}) \right|^2 \right] \\ &= \mathbb{E} \left[ \left| \frac{1}{(2Q+1)^2} \sum_{l=-Q}^Q \sum_{k=-Q}^Q [Y^2(\mathbf{x} + (l, k)\Delta) - \sigma_I^2(\mathbf{x} + (l, k)\Delta) - (\sigma_I^2(\mathbf{x}) - \sigma_I^2(\mathbf{x} + (l, k)\Delta))] \right|^2 \right] \\ &= \frac{1}{(2Q+1)^4} \mathbb{E} \left[ \left| \sum_{l=-Q}^Q \sum_{k=-Q}^Q [Y^2(\mathbf{x} + (l, k)\Delta) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)] \right|^2 \right] \end{aligned} \quad (18)$$



$$- \frac{2}{(2Q+1)^4} \mathbb{E} \left[ \left( \sum_{l=-Q}^Q \sum_{k=-Q}^Q (Y^2(\mathbf{x} + (l, k)\Delta) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)) \right) \sum_{l=-Q}^Q \sum_{k=-Q}^Q [\sigma_I^2(\mathbf{x}) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)] \right] \quad (19)$$

$$+ \frac{1}{(2Q+1)^4} \mathbb{E} \left[ \left| \sum_{l=-Q}^Q \sum_{k=-Q}^Q [\sigma_I^2(\mathbf{x}) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)] \right|^2 \right]. \quad (20)$$

When we simplify term (18), we have:

$$\begin{aligned} & \frac{1}{(2Q+1)^4} \mathbb{E} \left[ \sum_{\substack{l', k', l, k \\ = -Q}}^Q [Y^2(\mathbf{x} + (l, k)\Delta) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)] [Y^2(\mathbf{x} + (l', k')\Delta) - \sigma_I^2(\mathbf{x} + (l', k')\Delta)] \right] \\ &= \frac{1}{(2Q+1)^4} \sum_{\substack{l', k', l, k \\ = -Q}}^Q \mathbb{E} [\mathbb{E} [Y^2(\mathbf{x} + (l, k)\Delta) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)] [Y^2(\mathbf{x} + (l', k')\Delta) - \sigma_I^2(\mathbf{x} + (l', k')\Delta)] | \mathcal{F}^\sigma]] \\ &= \frac{1}{(2Q+1)^4} \sum_{\substack{l', k', l, k \\ = -Q}}^Q \mathbb{E} [\text{Cov} (Y^2(\mathbf{x} + (l, k)\Delta), Y^2(\mathbf{x} + (l', k')\Delta) | \mathcal{F}^\sigma)] \\ &= \frac{2 \sum_{\substack{l', k', l, k \\ = -Q}}^Q \mathbb{E} \left[ \left( \int_{\mathbb{R}^2} g(\mathbf{x} + (l, k)\Delta - \boldsymbol{\xi}) g(\mathbf{x} + (l', k')\Delta - \boldsymbol{\xi}) \sigma^2(\boldsymbol{\xi}) d\boldsymbol{\xi} \right)^2 \right]}{(2Q+1)^4} \rightarrow 0, \end{aligned} \quad (21)$$

by our assumption and Corollary 2.

Since  $\mathbb{E} [Y^2(\mathbf{x} + (l, k)\Delta) | \mathcal{F}^\sigma] = \sigma_I^2(\mathbf{x} + (l, k)\Delta)$  for  $-Q \leq l, k \leq Q$ , we also find that term (19) is equal to:

$$- \frac{2}{(2Q+1)^4} \mathbb{E} \left[ \left( \sum_{l=-Q}^Q \sum_{k=-Q}^Q [\sigma_I^2(\mathbf{x}) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)] \right) \sum_{l=-Q}^Q \sum_{k=-Q}^Q \mathbb{E} [(Y^2(\mathbf{x} + (l, k)\Delta) - \sigma_I^2(\mathbf{x} + (l, k)\Delta)) | \mathcal{F}^\sigma] \right],$$

which vanishes to zero. Simplifying term (20), we get:

$$\frac{\sum_{\substack{l', k', l, k \\ = -Q}}^Q [\mathbb{E} [\sigma_I^4(\mathbf{x})] - \mathbb{E} [\sigma_I^2(\mathbf{x}) \sigma_I^2(\mathbf{x} + (l, k)\Delta)] - \mathbb{E} [\sigma_I^2(\mathbf{x}) \sigma_I^2(\mathbf{x} + (l', k')\Delta)] + \mathbb{E} [\sigma_I^2(\mathbf{x} + (l, k)\Delta) \sigma_I^2(\mathbf{x} + (l', k')\Delta)]]}{(2Q+1)^4}$$

$$\begin{aligned} &= \frac{1}{(2Q+1)^4} \sum_{\substack{l', k', l, k \\ = -Q}}^Q [\text{Var} [\sigma_I^2(\mathbf{x})] - \text{Cov} (\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l, k)\Delta)) - \text{Cov} (\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l', k')\Delta)) \\ &+ \text{Cov} (\sigma_I^2(\mathbf{x} + (l, k)\Delta), \sigma_I^2(\mathbf{x} + (l', k')\Delta))] \text{ since } \sigma_I^2 \text{ is stationary,} \\ &= \frac{1}{(2Q+1)^4} \sum_{\substack{l', k', l, k \\ = -Q}}^Q [\text{Var} [\sigma_I^2(\mathbf{x})] - \text{Cov} (\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l, k)\Delta))] \end{aligned} \quad (22)$$

$$+ \frac{1}{(2Q+1)^4} \sum_{\substack{l', k', l, k \\ = -Q}}^Q [\text{Cov} (\sigma_I^2(\mathbf{x} + (l, k)\Delta), \sigma_I^2(\mathbf{x} + (l', k')\Delta)) - \text{Cov} (\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l', k')\Delta))] \quad (23)$$

Since we assumed that  $C(d_{x_1}, d_{x_2}) := \text{Cov} (\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (d_{x_1}, d_{x_2})))$  has a finite gradient over  $\mathbb{R}^2$ , i.e.  $\sigma_I^2$  has a second order mean squared derivative (page 27 of Adler (2010)), we can use the Mean Value Theorem and there exists

points  $\{c_{(l,k)} : -Q \leq l, k \leq Q\}$  in  $\mathbb{R}^2$  such that:

$$\begin{aligned}
& \left| \frac{1}{(2Q+1)^4} \sum_{\substack{l',k',l,k \\ =-Q}}^Q [\text{Var}[\sigma_I^2(\mathbf{x})] - \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l, k)\Delta))] \right| \\
& \leq \frac{1}{(2Q+1)^2} \sum_{l,k=-Q}^Q |\text{Var}[\sigma_I^2(\mathbf{x})] - \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l, k)\Delta))| \\
& \leq \frac{1}{(2Q+1)^2} \sum_{l,k=-Q}^Q |\nabla C(c_{(l,k)})| |(l, k)\Delta| \text{ where } \nabla C \text{ denotes the gradient of } C \text{ with respect to } d_{x_1} \text{ and } d_{x_2}, \\
& \leq \frac{K\Delta}{(2Q+1)^2} \sum_{l,k=-Q}^Q \sqrt{l^2 + k^2} \text{ where } K \text{ is the maximum absolute value of the gradients,} \\
& \leq \frac{K\Delta}{(2Q+1)^2} \sum_{l,k=-Q}^Q (|l| + |k|) \text{ by the Triangle Inequality,} \\
& = \frac{2K\Delta Q(Q+1)}{(2Q+1)^2},
\end{aligned}$$

which converges to zero if  $Q$  tends to infinity and  $\Delta$  behaves like  $O(Q^{-\tilde{r}})$  for  $\tilde{r} > 0$ .

Similarly, for term (23), there exists points  $\{c_{(l,k,l',k')} : -Q \leq l, k, l', k' \leq Q\}$  such that:

$$\begin{aligned}
& \left| \frac{1}{(2Q+1)^4} \sum_{\substack{l',k',l,k \\ =-Q}}^Q [\text{Cov}(\sigma_I^2(\mathbf{x} + (l, k)\Delta), \sigma_I^2(\mathbf{x} + (l', k')\Delta)) - \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l', k')\Delta))] \right| \\
& \leq \frac{1}{(2Q+1)^4} \sum_{\substack{l',k',l,k \\ =-Q}}^Q |\text{Cov}(\sigma_I^2(\mathbf{x} + (l, k)\Delta), \sigma_I^2(\mathbf{x} + (l', k')\Delta)) - \text{Cov}(\sigma_I^2(\mathbf{x}), \sigma_I^2(\mathbf{x} + (l', k')\Delta))| \\
& \leq \frac{1}{(2Q+1)^4} \sum_{\substack{l',k',l,k \\ =-Q}}^Q |\nabla f(c_{(l,k,l',k')})| |(l, k)(\Delta)| \\
& \leq \frac{K'\Delta}{(2Q+1)^2} \sum_{l,k=-Q}^Q \sqrt{l^2 + k^2} \text{ where } K' \text{ is the maximum absolute value of the gradients,} \\
& = \frac{2K'\Delta Q(Q+1)}{(2Q+1)^2}.
\end{aligned}$$

So, term (23) also converges to zero if  $Q$  tends to infinity and  $\Delta$  behaves like  $O(Q^{-\tilde{r}})$  for  $\tilde{r} > 0$ . Since under the latter conditions the MSE of our local variance estimator decreases to zero, we have proved that it converges to the true local variance in the  $\mathcal{L}_2$  sense.  $\square$

*Proof of Corollary 4.* Since  $\mathcal{L}_2$  convergence implies convergence in probability, each point at which we compute  $\hat{\sigma}_I^2$  is associated with a sequence  $\{\hat{\sigma}_I^2(\mathbf{x}, Q)\}$  converges in probability towards  $\sigma_I^2(\mathbf{x})$ . By Theorem 2.7(vi) of Van der Vaart (1998), this means that the vector  $(\hat{\sigma}_I^2(\mathbf{x}_1, Q), \dots, \hat{\sigma}_I^2(\mathbf{x}_M, Q))$  converges in probability to  $(\sigma_I^2(\mathbf{x}_1), \dots, \sigma_I^2(\mathbf{x}_M))$ . Furthermore, since convergence in probability implies convergence in distribution (also established in Theorem 2.7(ii) of Van der Vaart (1998)), we have that the limit of the joint distribution of  $(\hat{\sigma}_I^2(\mathbf{x}_1, Q), \dots, \hat{\sigma}_I^2(\mathbf{x}_M, Q))$  is the joint distribution of  $(\sigma_I^2(\mathbf{x}_1), \dots, \sigma_I^2(\mathbf{x}_M))$ . This means that the mean, variance and normalised variogram of the local variance estimator converge to those of the true conditional variance.  $\square$

## Acknowledgements

We would like to thank Mikko Pakkanen, Claudia Klüppelberg and Carsten Chong for helpful discussions. M. Nguyen is grateful to Imperial College for her PhD scholarship which supported this research. A.E.D. Veraart acknowledges financial support by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Programme.

## References

- Adler, R. J. (2010), *The geometry of random fields*, Vol. 62, SIAM.
- Applebaum, D. (2009), *Lévy processes and stochastic calculus*, Cambridge university press.
- Barndorff-Nielsen, O. E., Benth, F. E. & Veraart, A. E. D. (2015), Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency, in 'Banach Center Publications', Vol. 104, pp. 25–60.
- Barnes, S. L. (1964), 'A technique for maximizing details in numerical weather map analysis', *Journal of Applied Meteorology* **3**(4), 396–409.
- Bennedsen, M., Lunde, A. & Pakkanen, M. S. (2015), Hybrid scheme for Brownian semistationary processes, Technical Report 43, CREATES.
- Bolin, D. (2014), 'Spatial Matérn Fields Driven by Non-Gaussian Noise', *Scandinavian Journal of Statistics* **41**(3), 557–579.
- Cressie, N. (1993), *Statistics for spatial data*, John Wiley & Sons.
- Fouedjio, F., Desassis, N. & Rivoirard, J. (2016), 'A generalized convolution model and estimation for non-stationary random functions', *Spatial Statistics* **16**, 35–52.
- Fuentes, M. & Smith, R. L. (2001), A new class of nonstationary spatial models, Technical report, North Carolina State University, Raleigh, NC.
- Getis, A. (2015), 'Analytically derived neighborhoods in a rapidly growing West African city: The case of Accra, Ghana', *Habitat international* **45**, 126–134.
- Higdon, D. (1998), 'A process-convolution approach to modelling temperatures in the North Atlantic Ocean', *Environmental and Ecological Statistics* **5**(2), 173–190.
- Huang, W., Wang, K., Breidt, F. J. & Davis, R. A. (2011), 'A class of stochastic volatility models for environmental applications', *Journal of Time Series Analysis* **32**(4), 364–377.
- International Research Institute for Climate and Society (2016), 'Weekly sea surface temperature anomaly'. Available at [http://iridl.ldeo.columbia.edu/maproom/Global/Ocean\\_Temp/Weekly\\_Anomaly.html](http://iridl.ldeo.columbia.edu/maproom/Global/Ocean_Temp/Weekly_Anomaly.html) [Accessed on 7 June 2016].
- Johannesson, P., Podgórski, K., Rychlik, I. & Shariati, N. (2016), 'AR(1) time series with autoregressive gamma variance for road topography modeling', *Probabilistic Engineering Mechanics* pp. 106–116.
- Lahiri, S., Lee, Y. & Cressie, N. (2002), 'On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters', *J. Statist. Plann. Inference* **103**(1), 65–85.
- Pakkanen, M. S. (2014), 'Limit theorems for power variations of ambit fields driven by white noise', *Stochastic Processes and their Applications* **124**(5), 1942–1973.
- Rajput, B. S. & Rosinski, J. (1989), 'Spectral representations of infinitely divisible processes', *Probability Theory and Related Fields* **82**(3), 451–487.

- Sato, K.-I. (2007), 'Transformations of infinitely divisible distributions via improper stochastic integrals', *Alea Latin American Journal of Probability and Mathematical Sciences* **3**, 67–110.
- Seekell, D. A. & Dakos, V. (2015), 'Heteroskedasticity as a leading indicator of desertification in spatially explicit data', *Ecology and Evolution* .
- Van der Vaart, A. W. (1998), *Asymptotic statistics*, Cambridge University Press.
- Veraart, A. E. D. (2015), 'Stationary and multi-self-similar random fields with stochastic volatility', *Stochastics* **87**(5), 848–870.
- Wallin, J. & Bolin, D. (2015), 'Geostatistical Modelling Using Non-Gaussian Matérn Fields', *Scandinavian Journal of Statistics* **42**(3), 872–890.
- Yan, J. (2007), 'Spatial stochastic volatility for lattice data', *Journal of agricultural, biological, and environmental statistics* **12**(1), 25–40.
- Michele Nguyen, Department of Mathematics, Imperial College London, 180 Queen's Gate, SW7 2AZ London, UK.  
Email: michele.nguyen09@imperial.ac.uk